16 Combinatorial and parallel synthesis

Combinatorial and parallel synthesis have become established tools in drug discovery and drug development, allowing the use of a defined reaction route to produce a large number of compounds in a short period of time. The full set of compounds produced in this way is called a **compound library**. Reactions are usually carried out on small scale and the process can be automated or semi-automated, allowing reactions to be carried out in several reaction vessels at the same time and under identical conditions, but using different reagents for each vessel. The compact nature of the apparatus means that the process can be carried out within a normal fume cupboard.

16.1 Combinatorial and parallel synthesis in medicinal chemistry projects

In the past, medicinal chemistry involved the identification of a lead compound having a useful activity which was then modified to develop a clinically useful drug. Identification of the molecular target for the drug, and the mechanism by which it worked, often took many years to establish. Today, most medicinal chemistry projects start with an identifiable target, and the emphasis is on discovering a lead compound that will interact with this target. This reversal of priorities came about as a result of the human genome project and the proteomic revolution that followed. Once the genome was mapped, a vast number of previously unknown proteins were identified, all of which could be considered as potential drug targets. Pharmaceutical companies were faced with the problem of identifying the function of each target and finding a lead compound to interact with it. Before the advent of combinatorial chemistry and parallel synthesis, the need to find a lead compound was becoming the limiting factor in the whole process. Now, with the aid of these techniques, research groups can rapidly synthesize and screen thousands of structures in order to find new lead compounds, identify structure–activity relationships, and find analogues with good activity and minimal side effects (Fig. 16.1).

The procedures used in combinatorial synthesis are designed to produce mixtures of different compounds within each reaction vessel, whereas those used in parallel synthesis produce a single product in each vessel. In general, parallel synthesis is favoured because it is easier to identify the structures that are synthesized. However, there is still scope for combinatorial chemistry in finding lead compounds, especially as this procedure can generate significantly more structures in a set period of time, thus increasing the chances of finding a lead compound. Both methods generally involve the use of solid phase techniques, which are discussed in the next section.

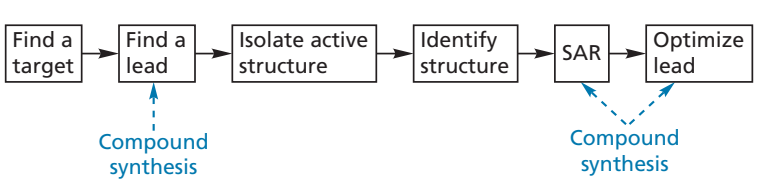


FIGURE 16.1 The stages in drug development requiring synthesis of large numbers of compounds (SAR = structure–activity relationships).

16.2 Solid phase techniques

Solid phase techniques can be used to carry out reactions where the starting material is linked to a solid support, such as a resin bead. Several reactions can then be carried out in sequence on the attached molecule. The final structure is then detached from the solid support. There are several advantages to this:

- since the starting material, intermediates, and final product are bound to a solid support, excess reagents or unbound by-products from each reaction can be easily removed by washing the resin;
- large excesses of reagents can be used to drive the reactions to completion (greater than 99%) because of the ease with which excess reagent can be removed;
- intermediates in a reaction sequence are bound to the bead and do not need to be purified;
- the polymeric support can be regenerated and reused if appropriate cleavage conditions and suitable anchor/linker groups are chosen (see later);
- automation is possible;
- if a combinatorial synthesis is being carried out, a range of different starting materials can be bound to separate beads. The beads can be mixed together such that all the starting materials are treated with another reagent in a single experiment. The starting materials and products are still physically distinct, as they are bound to separate beads. In most cases, mixing all the starting materials together in solution chemistry is a recipe for disaster, with polymerizations and side reactions producing a tarry mess. The individual beads can be separated at the end of the experiment to give individual products.

The essential requirements for solid phase synthesis are:

- a cross-linked insoluble polymeric support which is inert to the synthetic conditions (e.g. a resin bead);
- an anchor or linker covalently linked to the resin—the anchor has a reactive functional group that can be used to attach a substrate;
- a bond linking the substrate to the linker, which will be stable to the reaction conditions used in the synthesis;
- a means of cleaving the product or the intermediates from the linker;
- protecting groups for functional groups not involved in the synthetic route.

16.2.1 The solid support

The first successful example of solid phase synthesis was the **Merrifield resin** peptide synthesis. The resin involved consisted of polystyrene beads where the styrene is partially cross-linked with 1% divinylbenzene. The beads are derivatized with a chloromethyl group (the anchor/linker) to which amino acids can be coupled via an ester group (Fig. 16.2). This ester group is stable to the reaction

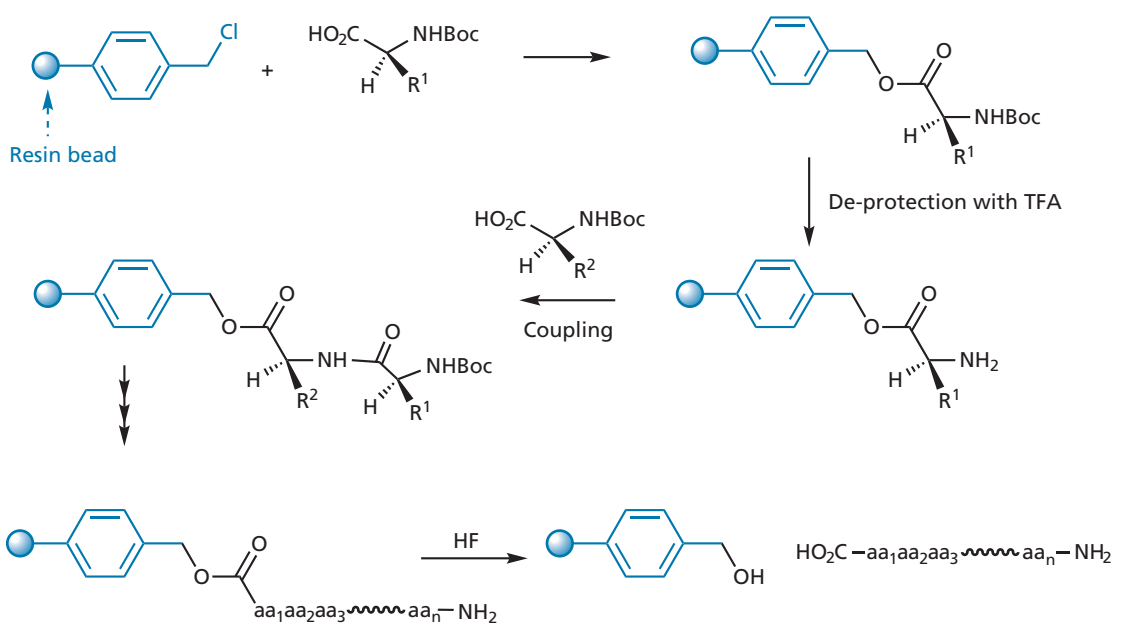


FIGURE 16.2 Peptide synthesis on a solid support (Boc = *tert*-butyloxycarbonyl (*t*-BuO-CO); TFA = trifluoroacetic acid).

conditions used in peptide synthesis, but can be cleaved at the end of the synthesis using vigorous acidic conditions (hydrofluoric acid).

One disadvantage of polystyrene beads is the fact that they are hydrophobic and the growing peptide chain is hydrophilic. As a result, the growing peptide chain is not solvated and often folds in on itself to form internal hydrogen bonds. This, in turn, hinders access of further amino acids to the exposed end of the growing chain. To address this, more polar solid phases were developed, such as **Sheppard's polyamide resin**. Other resins have been developed to be more suitable for the synthesis of non-peptides. For example, **Tentagel resin** is 80% polyethylene glycol grafted to cross-linked polystyrene, and provides an environment similar to ether or tetrahydrofuran. Regardless of the polymer that is used, the bead should be capable of swelling in solvent while remaining stable. Swelling is important because most of the reactions involved in solid phase synthesis take place in the interior of the bead rather than on the surface. It is wrong to think of resin beads as being like miniature marbles with an impenetrable surface. Each bead is a polymer and swelling involves unfolding of the polymer chains such that solvent and reagents can move between the chains into the heart of the polymer (Fig. 16.3).

Although beads are the common shape for the solid support, a range of other shapes, such as pins, have been designed to maximize the surface area available for reaction and, hence, maximize the amount of compound linked to the solid support. Functionalized glass surfaces have also been used and are suitable for oligonucleotide synthesis.

16.2.2 The anchor/linker

The anchor/linker is a molecular unit covalently attached to the polymer chain making up the solid support. It contains a reactive functional group with which the starting material in the proposed synthesis can react and hence become attached to the resin. The resulting link must be stable to the reaction conditions used throughout the synthesis, but be easily cleaved to release the final compound once the synthesis is complete (Fig. 16.4). As the linkers are distributed along the length of the polymer chain, most of them will be in the interior of the polymer bead, emphasizing the importance of the bead swelling if the starting material is to reach them.

Different linkers are used depending on:

- the functional group which will be present on the starting material;
- the functional group which is desired on the final product once it is released.

Resins having different linkers are given different names (Fig. 16.5). For example, the **Wang resin** has a linker which is suitable for the attachment and release of carboxylic acids. It can be used in peptide synthesis by linking an *N*-protected amino acid to the resin by means of an ester link. This ester link remains stable to coupling and de-protection steps in the peptide synthesis, and can then be cleaved using trifluoroacetic acid (TFA) to release the final peptide from the bead (Fig. 16.6). One problem with the Wang resin is that the first amino acid linked to the resin is prone to **racemization**. The **Barlos resin** contains a trityl linker and was designed to avoid this problem. The final product can be cleaved under very mild conditions (e.g. HOAc/TFE/CH_2Cl_2 or TFA/CH_2Cl_2) owing to the high stability of the trityl cations that are formed. Molecules can also be linked to the resin by means of an alcohol group.

Starting materials with a carboxylic acid (RCO_2H) can be linked to the **Rink resin** via an amide link. Once the reaction sequence is complete, treatment with TFA releases the product with a primary amide group, rather than the original carboxylic acid (**R′**$CONH_2$; Fig. 16.7).

Primary and secondary alcohols (ROH) can be linked to a dihydropyran-functionalized resin. Linking the alcohol is done in the presence of pyridinium 4-toluenesulfonate (PPts) in dichloromethane. Once the reaction sequence has been completed, cleavage can be carried out using TFA (Fig. 16.8).

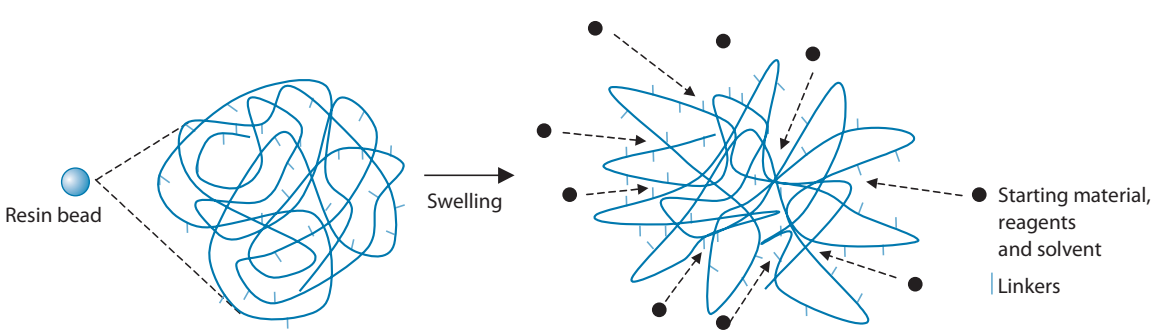


FIGURE 16.3 Swelling of a resin bead allowing access of reagents and solvent.

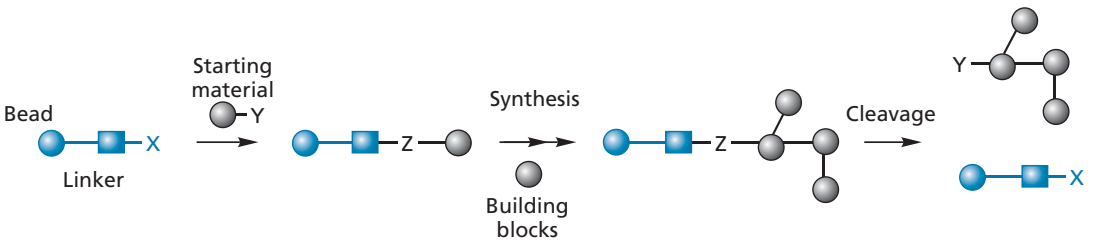


FIGURE 16.4 The principles of an anchor/linker. X, Y, Z are functional groups.

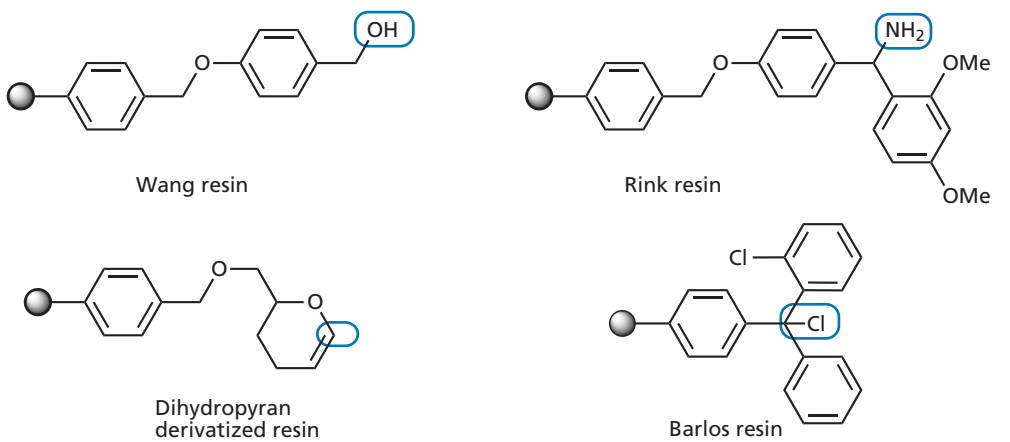


FIGURE 16.5 Types of resin with the linkage point circled.

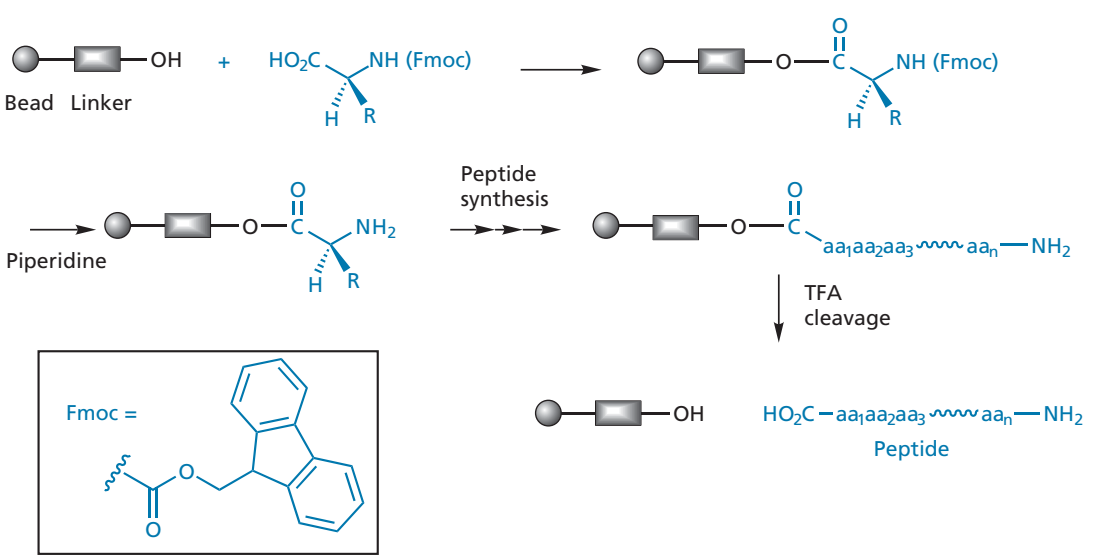


FIGURE 16.6 Peptide synthesis with a Wang resin—the structure of the linker is shown in Fig. 16.5.

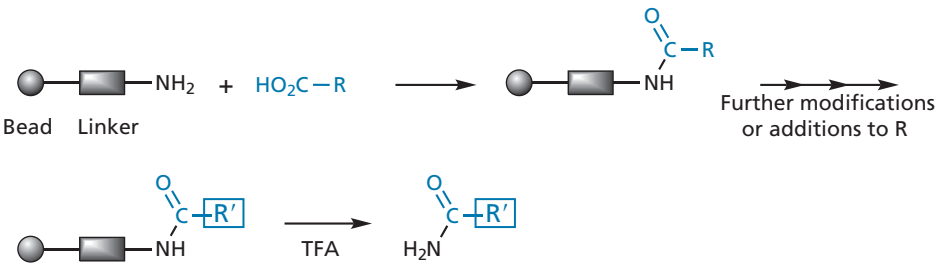


FIGURE 16.7 Solid phase synthesis with a Rink resin (R contains functional groups which allows further modifications of the molecule to give R'). The structure of the linker is shown in Fig. 16.5.

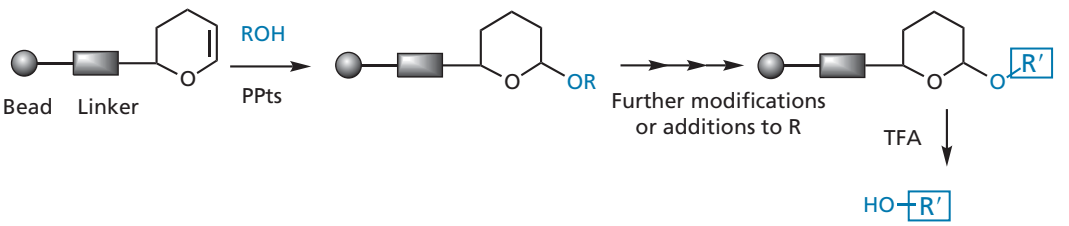


FIGURE 16.8 Solid phase synthesis with a dihydropyran-functionalized resin (R contains functional groups which allows further modifications of the molecule to give R').

16.2.3 Examples of solid phase syntheses

Solid phase synthesis was pioneered by Merrifield, and most of the early work involved peptide synthesis. However peptides pose particular problems as drugs in terms of their pharmacokinetic properties (section 14.8.2), and so a large amount of research was carried out to extend solid phase synthetic methods to the synthesis of small non-peptide molecules. The first move away from natural peptides was to use the same peptide coupling procedures, but with non-natural amino acids. Peptides could also be modified once they were built by reactions such as *N*-methylation. *N*-Substituted glycine units were used to produce structures known as **peptoids** where the side chain is attached to the nitrogen rather than the α-carbon. Some of these have been shown to be ligands for various important receptors and show increased metabolic stability.

A disadvantage with all the above structures is the fact that they are linear, flexible molecules linked together by a regular molecular backbone. The real interest in solid phase synthesis began when it became possible to produce heterocyclic structures. Heterocycles are less susceptible to metabolism, and have better pharmacokinetic properties. They are more rigid, and diversity is possible by varying the substituents around the heterocyclic 'core'.

1,4-Benzodiazepines have been synthesised by linking a selection of amino acids to resin beads through the carboxylic acid group (Fig. 16.9). Reaction with a variety of imines gave the adducts shown. Treatment with TFA

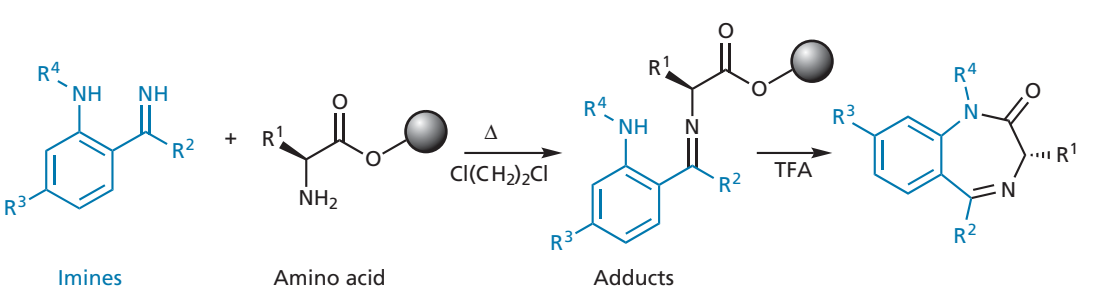


FIGURE 16.9 Benzodiazepine synthesis involving a cyclo-release strategy.

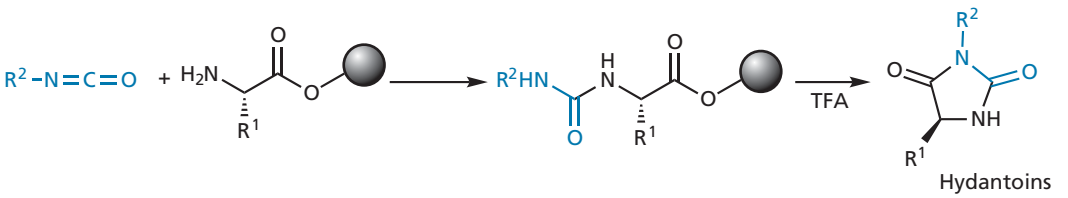


FIGURE 16.10 Synthesis of hydantoins.

released the adducts which then cyclized to give the final products. The advantage of this synthesis lies in the fact that the functional group released from the resin takes part in the final cyclization and does not remain as an extra, and possibly redundant, group. The final product has four variable substituents spread evenly around the bicyclic ring system. This allows exploration of conformational space around the whole molecule when searching for binding interactions with a drug target (see section 16.3.1).

A similar strategy was employed for the synthesis of hydantoins (Fig. 16.10), and a large variety of heterocyclic compounds have now been synthesized using solid phase methods.

The range of reactions which can be carried out on solid phase has also been extended: most common reactions are now feasible, including moisture sensitive and organometallic reactions. For example, aldol condensations, DIBAL reductions, Wittig reactions, LDA reductions, Heck couplings, Stille couplings, and Mitsunobu reactions are all possible. Automated or semi-automated synthesizers can cope with 6, 12, 42, 96, or 144 reaction vials depending on the instrument and the size of the reaction tubes used. The addition of solvent, starting materials, and reagents can be carried out automatically using syringes. Automated work-up procedures, such as the removal of solvent, washing and liquid–liquid separations are also possible. Reactions can be stirred and carried out under inert atmospheres, and the reactions can be heated or cooled as required.

16.3 Planning and designing a compound library

The techniques of solid phase synthesis have been used to produce large quantities of compounds from a particular reaction sequence. These can be stored as compound libraries and then accessed to search for new lead compounds capable of interacting with novel or existing drug targets. It is important that the molecules in these libraries are structurally diverse to increase the chances of success, and so some thought has to be put into planning and designing a compound library.

16.3.1 'Spider-like' scaffolds

A compound library is generated from a specific sequence of reactions, so there is a danger that it will contain a large number of very similar molecules. Therefore, care has to be taken about the type of molecule synthesized, the synthetic route employed, and the types of substituents involved, in order to achieve structural diversity. In general, it is best to synthesize 'spider-like' molecules (**spider scaffolds**), so-called because they consist of a central body (called the **centroid** or **scaffold**) from which various 'arms' (substituents) radiate (Fig. 16.11). These arms contain different functional groups which are used to probe a binding site for binding regions once the spider-like molecule has entered (Fig. 16.12). The chances of success are greater if the 'arms' are evenly spread around the scaffold, as this allows a more thorough exploration of the three-dimensional space (**conformational space**) around the molecule. The molecules made in the synthesis are planned in advance to ensure that they contain different functional groups on their arms, placed at different distances from the central scaffold.

16.3.2 Designing 'drug-like' molecules

The 'spider-like' approach increases the chances of finding a lead compound which will interact with a target

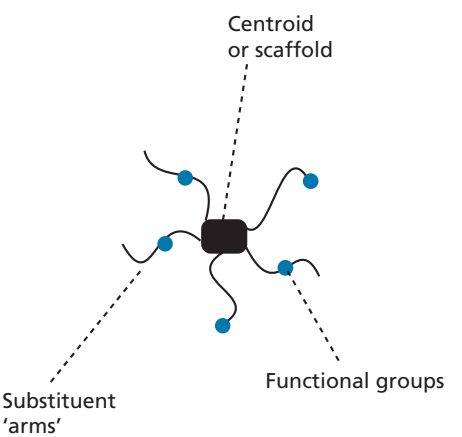


FIGURE 16.11 'Spider-like' molecule.

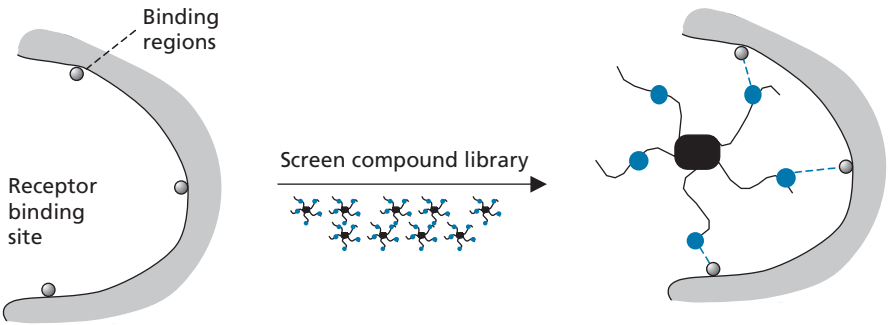


FIGURE 16.12 Probing for an interaction.

binding site, but it is also worth remembering that compounds with good binding interactions do not necessarily make good medicines. There are also the pharmacokinetic issues to be taken into account (Chapter 11), and so it is worthwhile introducing certain restrictions to the types of molecule that will be produced in order to increase the chances that the lead compound will be orally active. In general, the chances of oral activity are increased if the structure obeys Lipinski's rule of five or Veber's parameters (section 11.3). However, allowance has to be made for the fact that any lead compounds identified are almost certainly going to require substantial optimization, in which case more stringent guidelines should be applied (section 12.4.11). Other restrictions should be considered. For example, groups such as esters should be avoided because they are easily metabolized. Scaffolds or substituents likely to result in toxic compounds should also be avoided, for example alkylating groups or aromatic nitro groups.

16.3.3 Synthesis of scaffolds

Most scaffolds are constructed using the synthetic route employed for the solid phase synthesis, and this also determines the number and variety of substituents that can be attached to the scaffold. The ideal scaffold should be small in order to allow a wide variation of substituents (see Box 16.2). It should also have its substituents widely dispersed around its structure (spider-like) rather than restricted to one part of the structure (**tadpole-like**) if the conformational space around it is to be fully explored (Fig. 16.13). Finally, the synthesis should allow each of the substituents to be varied independently of each other.

Scaffolds can be flexible (e.g. a peptide backbone) or rigid (a cyclic system). They may contain heteroatoms that are capable of forming useful bonding interactions with the binding site or they may not. Some scaffolds are already common in medicinal chemistry (e.g. benzodiazepine, hydantoin, tetrahydroisoquinoline, benzenesulfonamide, and biaryls) and are associated with a diverse range of activities. Such scaffolds are termed **privileged scaffolds**.

16.3.4 Substituent variation

The variety of substituents chosen in a combinatorial synthesis depends on their availability and the diversity required. This would include such considerations as structure, size, shape, lipophilicity, dipole moment, electrostatic charge, and functional groups present. It is usually best to identify which of these factors should be diversified before commencing the synthesis.

16.3.5 Designing compound libraries for lead optimization

If a compound library is being planned for drug optimization, the variations planned should take into account several factors such as the biological and physical properties of the compound, its binding interactions with the target, and the potential problems of particular substituents. For example, if the binding interactions of a target receptor with its usual ligand are known, this knowledge can be used to determine what size of compounds would be best to synthesize, the types of functional groups

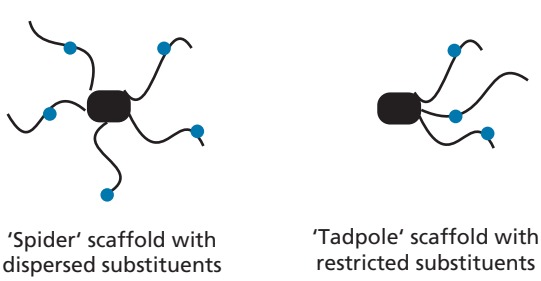


FIGURE 16.13 Dispersed and restricted substituents.

BOX 16.1 Examples of scaffolds

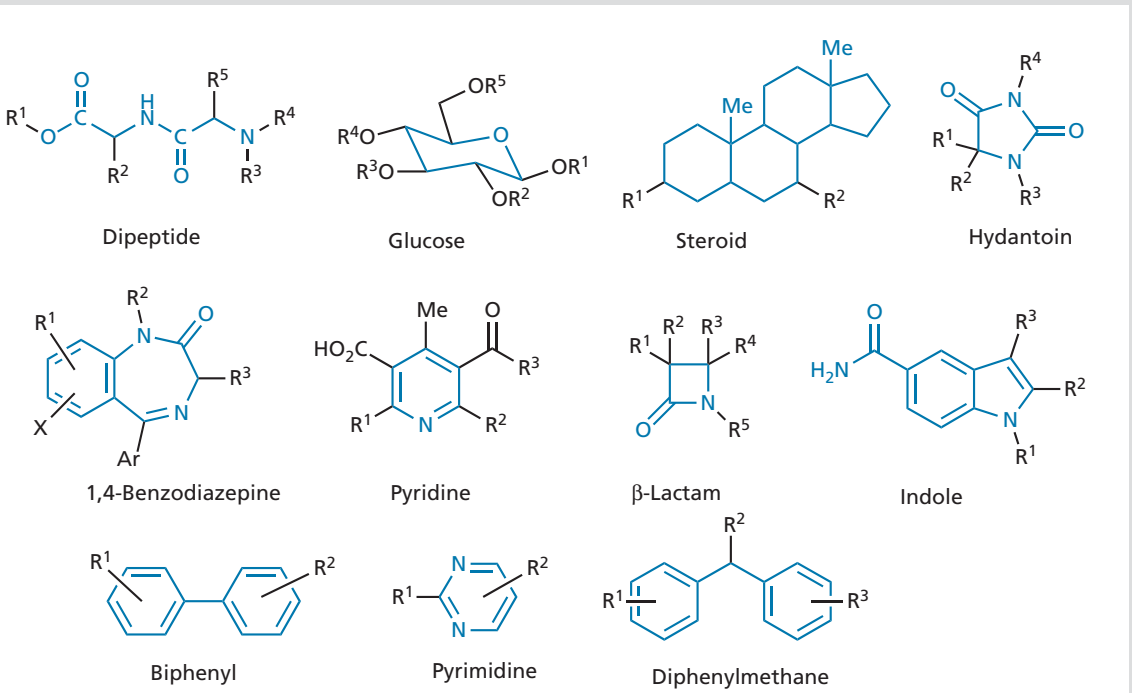
Benzodiazepines, hydantoins, β-lactams, and pyridines are examples of extremely good scaffolds. They all have small molecular weights, and there are various synthetic routes available which produce the substitution patterns required to fully explore the conformational space about them. For example, it is possible to synthesize benzodiazepines such that there are variable substituents round the whole structure.

Peptide scaffolds are flexible scaffolds which have the capacity to form hydrogen bonds with target binding sites. They are easy to synthesize and a large variety of different substituents are possible by using the amino acid building blocks. Further substitution is possible on the terminal amino and carboxylic acid functions. The substituents are widely distributed along the peptide chain allowing a good exploration of conformational space. If we consider Lipinski's rule of five, the peptide scaffold should, ideally, be restricted to di- and tripeptides in order to keep the molecular weight below 500. It is interesting to note that the orally active antihypertensive agents **captopril** and **enalapril** are dipeptide-like, whereas larger peptides, such as the **enkephalins**, are not orally active. Oral activity has also been a problem with HIV protease inhibitors that have molecular weights over 500 (section 20.7.4).

Some of the scaffolds shown in the diagram have various disadvantages. Although **glucose** has a small molecular weight and the possibility of five variable substituents, it contains multiple hydroxyl groups. Attaching different substituents to similar groups usually requires complex protection and de-protection strategies. Nevertheless, the potential of sugar-based drugs is so great that a lot of progress has been made in developing solid phase syntheses based on sugar scaffolds.

Steroids might appear attractive as scaffolds. However, the molecular weight of the steroid skeleton itself (314) limits the size and number of the substituents which can be added if we wish to keep the overall molecular weight below 500.

The indole scaffold shown suffers a disadvantage in having its variable substituents located in the same region of the molecule, preventing a full exploration of conformational space (i.e. it is a tadpole-like scaffold).



Examples of scaffolds.

that ought to be present, and their relative positions. For example, if the target is a zinc-containing protease (e.g. angiotensin-converting enzyme), a library of compounds containing a carboxylic acid or thiol group would be relevant.

16.3.6 Computer-designed libraries

It has been claimed that half of all known drugs involve only 32 scaffolds. Furthermore, it has been stated that a relatively small number of moieties account for the large

majority of side chains in known drugs. This may imply that it is possible to define 'drug-like molecules' and use computer software programs to design more focused compound libraries. Descriptors used in this approach include log *P*, molecular weight, number of hydrogen bond donors, number of hydrogen bond acceptors, number of rotatable bonds, aromatic density, the degree of branching in the structure, and the presence or absence of specific functional groups. One can also choose to filter out compounds that do not obey the rules mentioned in section 16.3.2. Computer programs can also be used to identify the structures which should be synthesized in order to maximize the number of different pharmacophores produced (section 17.16).

16.4 Testing for activity

We shall now look in more detail at how the structures in a compound library are tested for biological activity.

16.4.1 High-throughput screening

Because solid phase syntheses can produce a large quantity of structures in a very short period of time, biological testing has to be carried out quickly and automatically. The process is known as high-throughput screening (HTS) and was developed before combinatorial and parallel synthetic methods were devised. Indeed, the existence of HTS was one of the driving forces that led to the development of these synthetic procedures. Since biological testing was so rapid and efficient, the pharmaceutical companies soon ran out of novel structures to test, and the synthesis of new structures became the limiting factor in the whole process of drug discovery. Combinatorial and parallel syntheses have solved that problem and the number of new compounds synthesized each year has increased dramatically. In fact, there are now so many compounds being produced that the focus is on making HTS even more efficient. Traditionally, compounds are tested automatically and analysed on a plate containing 96 small wells, each with a capacity of 0.1 ml. There is now a move to use test plates of a similar size but containing 1536 wells, where the test volumes are only 1–10 µl. Moreover, methods such as fluorescence and chemiluminescence are being developed which will allow the simultaneous identification of active wells. Further miniaturization of open systems is unlikely because of the problems of evaporation involving volumes less than 1 µl. However, miniaturization using closed systems is on the horizon. The next major advance will be in the science of **microfluidics**, which involves the manipulation of tiny volumes of liquids in confined space. Microfluidic circuits on a chip can be used to control fluids electronically allowing separation of an analytical sample using capillary electrophoresis. Companies are now developing machines that combine ultra-small-scale synthesis (section 16.5.5) and miniaturized analysis. A single 10×10 cm silicon wafer can be microfabricated to support 10^5 separate syntheses/bioassays on a nanolitre scale!

16.4.2 Screening 'on bead' or 'off bead'

Sometimes structures can be tested for biological activity when they are still attached to the solid phase. 'On bead' screening assays involve interactions with targets which are tagged with an enzyme, fluorescent probe, radionuclide, or chromophore. A positive interaction results in a recognizable effect, such as fluorescence or a colour change. These screening assays are rapid and 10^8 beads can be readily screened. Active beads can then be picked out by micromanipulation and the structure of the active compound determined.

A false-negative might be obtained if the solid phase interferes sterically with the assay. If such interference is suspected, it is better to release the drug from the solid phase before testing. This avoids the uncertainty of false-negatives. However, there are cases where the compounds released prove to be insoluble in the test assay and give a negative result, whereas they give a positive result when attached to the bead.

KEY POINTS

- Solid phase synthetic methods have proved valuable in producing compounds for lead discovery, structure–activity relationships and drug optimization.
- Parallel synthesis involves the synthesis of a different compound in each reaction vial, and is useful in all aspects of medicinal chemistry where synthesis is required.
- Combinatorial synthesis involves the synthesis of mixtures of compounds in each reaction vial, and has been useful in discovering lead compounds.
- Solid phase synthesis has several advantages. Intermediates do not need to be isolated or purified. Reactants and reagents can be used in excess to drive the reaction to completion. Impurities and excess reagents or reactants are easily removed.
- In combinatorial syntheses different compounds are linked to different solid phase surfaces such that they are physically separated, allowing them to undergo reactions and work-up procedures in the same reaction vessel.
- The solid support consists of a polymeric surface and a linker molecule which allows a starting material to be linked covalently to the support.

- Different linkers are used depending on the functional group present on the starting material and the functional group that is desired on the product.
- A scaffold is the core structure of a molecule round which variations are possible through the use of different substituents.
- Spider-like scaffolds allow substituent variation around the whole molecule, making it possible to explore all the conformational space around the scaffold. This increases the possibility of finding a lead compound which will bind to a target binding site.
- Lipinski's rule of five can be used when planning compound libraries to increase the chances of identifying an orally active lead compound. More rigid guidelines may be used if the lead compound is likely to undergo substantial optimization.
- A privileged scaffold is one which is commonly present in known drugs.
- Computer software is available to assist in the planning of compound libraries
- High-throughput screening allows the automated analysis of large numbers of samples for their biological activity against defined targets. The analysis requires only small quantities of each sample.
- Screening can be carried out on compounds attached to resin beads, or on compounds which have been released into solution.

16.5 Parallel synthesis

In parallel synthesis, a reaction is carried out in a series of wells such that each well contains a single product. This method is a 'quality rather than quantity' approach and is often used for focused lead optimization studies. For parallel synthesis to be fast and efficient, it is necessary to remove or simplify the bottlenecks associated with classical organic synthesis. These include laborious work-ups, extractions, solvent evaporations, and purifications. A typical medicinal chemist may synthesize one or two new entities a week using the classical approaches to organic synthesis. With parallel synthesis, that same researcher can synthesize a dozen or more pure molecules, thus increasing the synthetic output and speeding up the lead optimization process. Parallel synthesis can be carried out on solid phase and we have already seen the advantages of solid phase synthesis (section 16.2). However, parallel syntheses can also be carried out in solution and in this section we focus on methods that make **solution phase organic synthesis** (**SPOS**) more efficient.

This can be illustrated by considering the synthesis of an amide, which typically involves the reaction of a carboxylic acid with an amine in the presence of a coupling reagent such as dicyclohexylcarbodiimide (DCC) (Fig. 16.14). Conventionally, a work-up procedure has to be carried out once the reaction is complete. This involves washing the organic solution with aqueous acid in order to remove unreacted amine. Once the aqueous and organic layers have been separated, the organic layer is washed with an aqueous base in order to remove unreacted acid. The organic and basic layers are separated, and then the organic layer is treated with a drying agent such as magnesium sulphate. The drying agent is filtered off and then the solvent is removed to afford the crude amide. Purification then has to be carried out by crystallization or chromatography. In order to synthesize a small, 12-component amide library by reacting different carboxylic acids with the same amine, one would have to repeat all of these steps and this would prove both time consuming and equipment intensive.

Equipment miniaturization for parallel synthesis means that it is possible to house a mini-parallel synthesis laboratory in a fume cupboard for each chemist (Fig. 16.15). Small footprint work stations often enable one to perform up to 24 reactions followed by 24 simultaneous evaporations on a normal heater stirrer unit. Multiple parallel or sequential automated chromatography units can facilitate purification, and microwave reactors can dramatically speed up reaction times. In this way, all 12 amides in our library can be made at the same time in parallel. A variety of useful techniques can also be used to minimize the work-up procedures as described in the next sections.

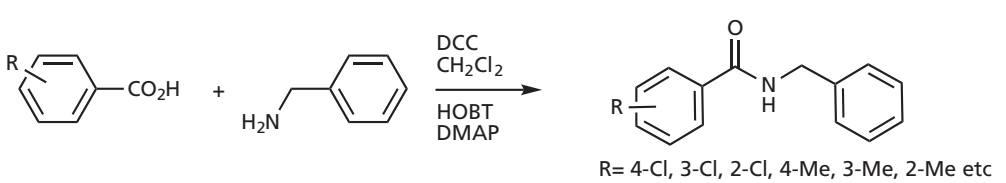


FIGURE 16.14 Coupling reaction of a carboxylic acid and an amine to give an amide library (DCC = dicyclohexylcarbodiimide; DMAP = dimethylaminopyridine; HOBT = hydroxybenztriazole).

FIGURE 16.15 Laboratory stations for microwave-assisted organic reactions (CEM Explorer-24) and parallel synthesis (Radley's Greenhouse).

16.5.1 Solid phase extraction

Solid phase extraction (SPE) can be used to avoid the 'hassle' of carrying out liquid–liquid extractions to remove acidic or basic compounds, or impurities. For example, solutions of the 12 crude amides that have been prepared can be taken up from their wells at the same time using a multi-channel pipettor and applied to a battery of silica columns. An acidic column (SCX column) removes basic impurities, while a basic column (SAX column) removes acid impurities. Once the solutions have passed through the columns, the solvents are concentrated in parallel to yield essentially pure amide.

Another method of removing excess amine from a reaction is to use **fluorous solid phase extraction** (**F-SPE**). This consists of silica columns where the silica has been linked to alkyl chains containing a large number of fluoro substituents. The highly fluorinated silica has a high affinity for fluorinated molecules and can be used to separate fluorinated compounds from non-fluorinated compounds. For example, consider the reaction shown in Fig. 16.16, where an isocyanate is treated with an amine to give a urea product. The amine is used in excess in order to drive the reaction to completion, but the amine left over has to be removed. In order to do this, a fluorinated isocyanate is added which reacts with the excess amine to produce a fluorinated urea (Fig. 16.17). The crude solution is passed through an F-SPE column which acts as a scavenger resin to retain the highly fluorinated urea and allows the desired unfluorinated urea product to pass through.

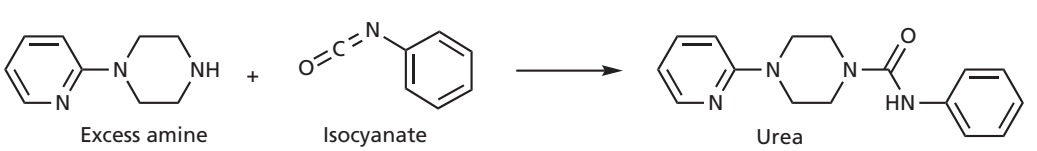


FIGURE 16.16 Reaction of an isocyanate with excess amine to produce a urea.

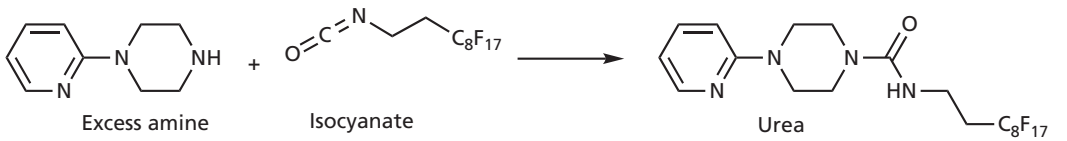


FIGURE 16.17 Removal of excess amine by reaction with a fluorinated isocyanate followed by F-SPE.

Sometimes an aqueous work-up cannot be avoided. For example, an aqueous work-up is required following a Grignard reaction, which means that the aqueous and organic phases have to be separated. Fortunately, there are efficient methods of carrying out such separations in parallel.

One such method is to use a **lollipop phase separator**. A pin is inserted into a mixture of the two phases and the mixture is rapidly cooled in a dry ice/acetone bath at –78 °C. The aqueous phase freezes onto the pin to form a 'lollipop'. The pin and its lollipop can then be removed from the reaction vial, leaving the organic phase behind. Up to 96 such separations may be performed in parallel with specially designed units.

Another method is to use **phase separation columns**, which can be used to separate a dense chlorinated organic layer from an aqueous phase. The lower organic layer passes through a hydrophobic frit by gravity, whereas the upper aqueous layer is retained on the frit. It is important not to apply pressure, otherwise the aqueous phase may also be forced through the frit.

16.5.2 The use of resins in solution phase organic synthesis (SPOS)

By carrying out a parallel synthesis in solution, it is easy to monitor the reaction by ^{1}H NMR spectroscopy or by thin layer chromatography. Work-up procedures can be greatly simplified by the use of a variety of resins. As resins are solid-supported, there is little interaction between different types, allowing a variety of resins to be used in the same reaction. Thus, it is common to have a reaction cocktail which includes nucleophilic and electrophilic resins, or acidic and basic resins without any problems arising.

Reactions are carried out such that one of the reagents—usually the cheaper and more readily available—is used in excess in order to drive the reaction to completion (Fig. 16.18). The crude reaction mixture will comprise the product AB and excess starting material A. The crude mixture is treated with a solid-supported scavenger resin that is capable of reacting with the excess reagent (A). As a result, excess reagent becomes attached to the resin and can be removed by filtering the resin. Removal of the solvent then leaves the pure product (AB).

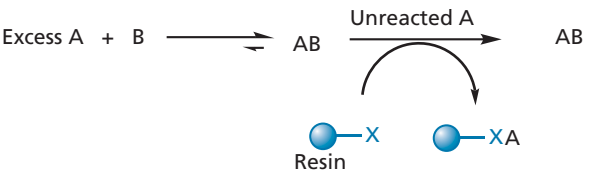


FIGURE 16.18 The use of scavenger resins in solution phase organic synthesis.

16.5.3 Reagents attached to solid support: catch and release

It is possible to attach a reagent to a solid support. This has the advantage that the reagent or its by-product can be removed easily at the end of the reaction. For example, the coupling agent used for amide synthesis can be attached to a resin instead of being present in solution (Fig. 16.19). The reaction involves a carboxylic acid starting material reacting with the coupling reagent to form an intermediate which is still linked to the resin. Thus, the carboxylic acid is taken out of solution—the 'catch' phase. The resin-bound intermediate now reacts with the amine and the amide product is released back into solution. The urea by-product which is formed remains bound to the resin and is easily removed when the resin is finally filtered. Acidic and basic resins can also be added to remove reagents and excess starting materials as described earlier.

The formation of a sulphonamide library shown in Fig. 16.20 makes use of a variety of different resins. The reaction involves an amine being treated with an excess of a sulphonyl chloride. A basic catalyst is required for the reaction and triethylamine is normally used in a conventional synthesis. However, this is quite a smelly, volatile compound and would have to be removed once the reaction was complete. Instead of triethylamine, a resin-bound base, such as morpholine (PS-morpholine), can be used.

Following the reaction, nucleophilic and electrophilic scavenger resins are added. The nucleophilic resin PS-trisamine reacts with excess sulphonyl chloride to remove it from solution, while the electrophilic resin PS-isocyanate removes unreacted amine (Fig. 16.21). Filtration, to remove the resins, leaves the pure sulphonamide in solution.

Solid supported reagents can be used in a variety of very common synthetic reactions. For example, a solid

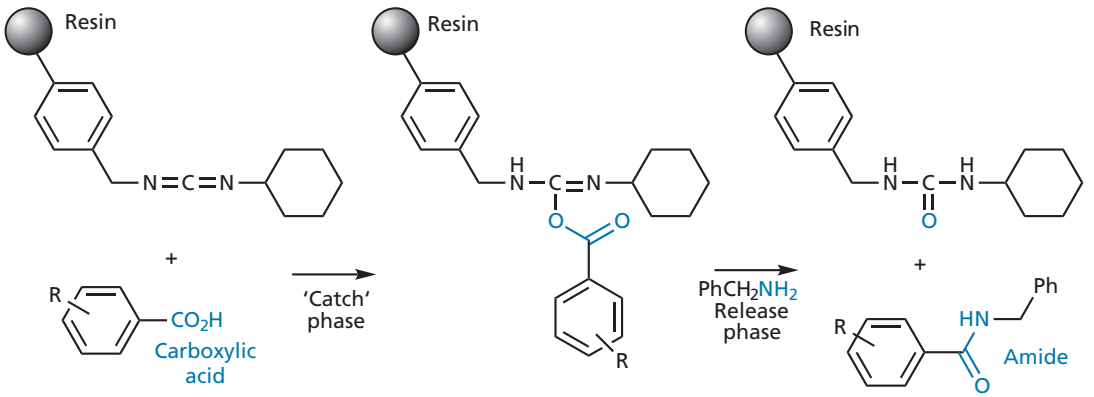


FIGURE 16.19 'Catch and release' during a coupling reaction.

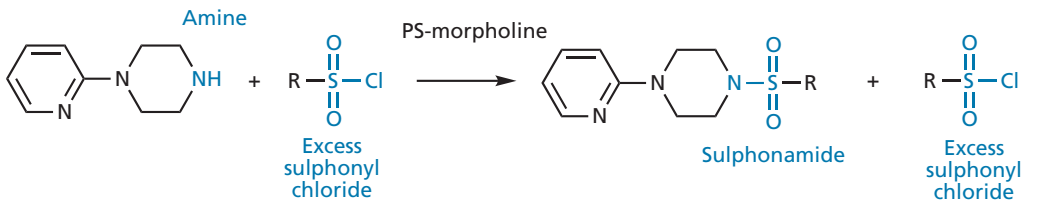


FIGURE 16.20 Formation of a sulphonamide library.

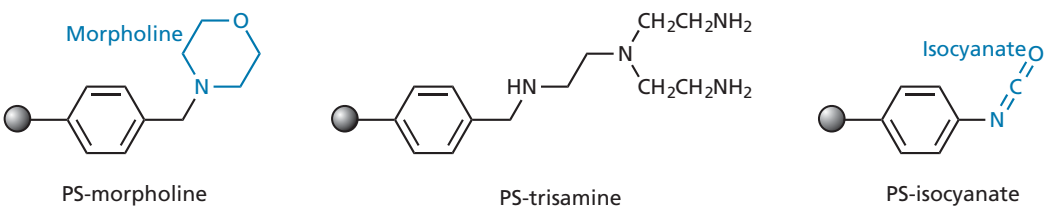


FIGURE 16.21 PS-morpholine, PS-trisamine, and PS-isocyanate.

supported borohydride can be used to reduce carbonyl groups (Fig. 16.22). In some reactions, it is also possible to reduce the toxicity and odour of reagents and their by-products. For example, the normal Swern oxidation involves the formation of dimethylsulphide as a by-product—a compound which has a pungent cabbage odour! This is avoided by using a solid supported reagent instead (Fig. 16.23).

16.5.4 Microwave technology

Drug discovery is a very expensive process and **microwave assisted organic synthesis** (MAOS) is proving to be a very useful tool for accelerating syntheses and making the process more efficient. There are many examples of thermal reactions that take several hours to complete using heaters or oil baths, but which are carried out in minutes using microwave conditions. There is a much greater efficiency of energy transfer using microwave technology which accounts for the faster reaction times. Moreover, yields can sometimes be improved dramatically with less decomposition and fewer side reactions. Specially designed microwave units are now commonly employed in library syntheses (Fig. 16.15). Examples of reactions that have been carried out using microwave technology include the formation of amides from acids and amines without the need for coupling agents (Fig. 16.24), metal-catalysed Suzuki couplings which can be performed even on usually unreactive aryl chlorides (Fig. 16.25), and metal-mediated reductions and aminations (Fig. 16.26). The reduction shown in Fig. 16.26 took 24 h using conventional heating and only 15 min using microwave heating.

16.5.5 Microfluidics in parallel synthesis

The science of **microfluidics** involves the manipulation of tiny volumes of liquids in a confined space. Companies are devising microreactors that can be used to carry out

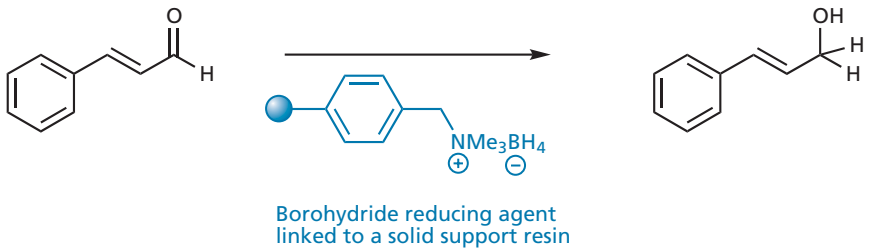


FIGURE 16.22 Reduction of an aldehyde with a solid-supported borohydride reagent.

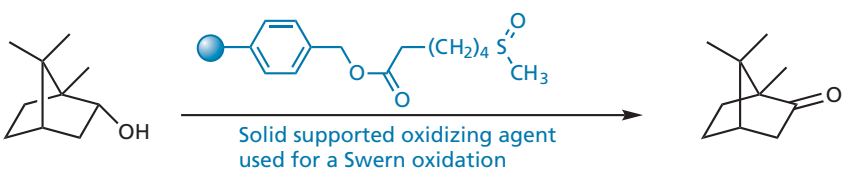


FIGURE 16.23 Swern oxidation using a solid supported reagent.

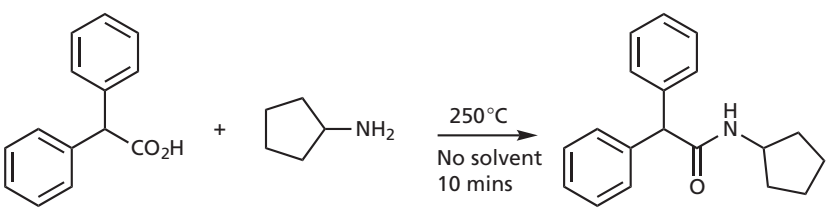


FIGURE 16.24 Amide formation using microwave technology.

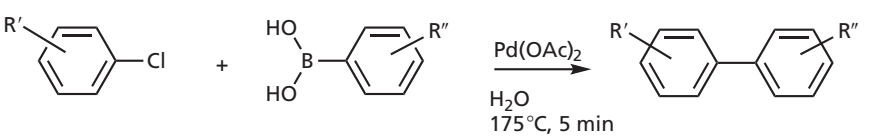


FIGURE 16.25 A Suzuki coupling carried out under microwave conditions.

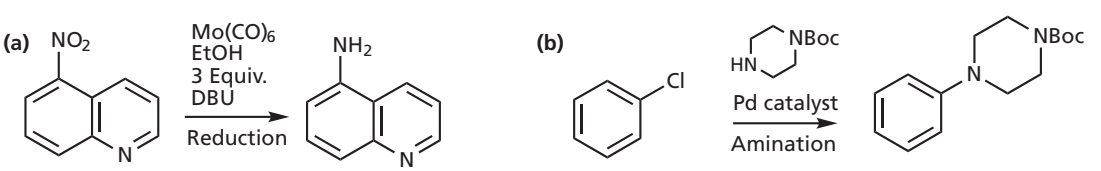


FIGURE 16.26 Microwave-assisted transition metal-mediated reactions. (**a**) Reductions and (**b**) aminations.

parallel syntheses on microchips (Fig. 16.27) using a continuous flow of reactants in microfluidic channels. The channels are designed such that various reactants are mixed and reacted as they flow through the microchip. Several reactions have already been carried out successfully at the microscale and it is found that many reaction times are shortened from hours to minutes. Some reactions take place in higher yield and with fewer side products. It is also possible to control the temperature of each reaction extremely accurately. Another advantage of microreactors is the potential to handle a vast number of parallel reactions on microchips. The channels through each chip can be fabricated to allow all possible mixing combinations of the various reactants, either on separate microchips or on a three-dimensional microchip. The example in Fig. 16.27 is a simple illustration of how a microreactor system could be set up to create a mini-library from the reaction of A or B with C or D.

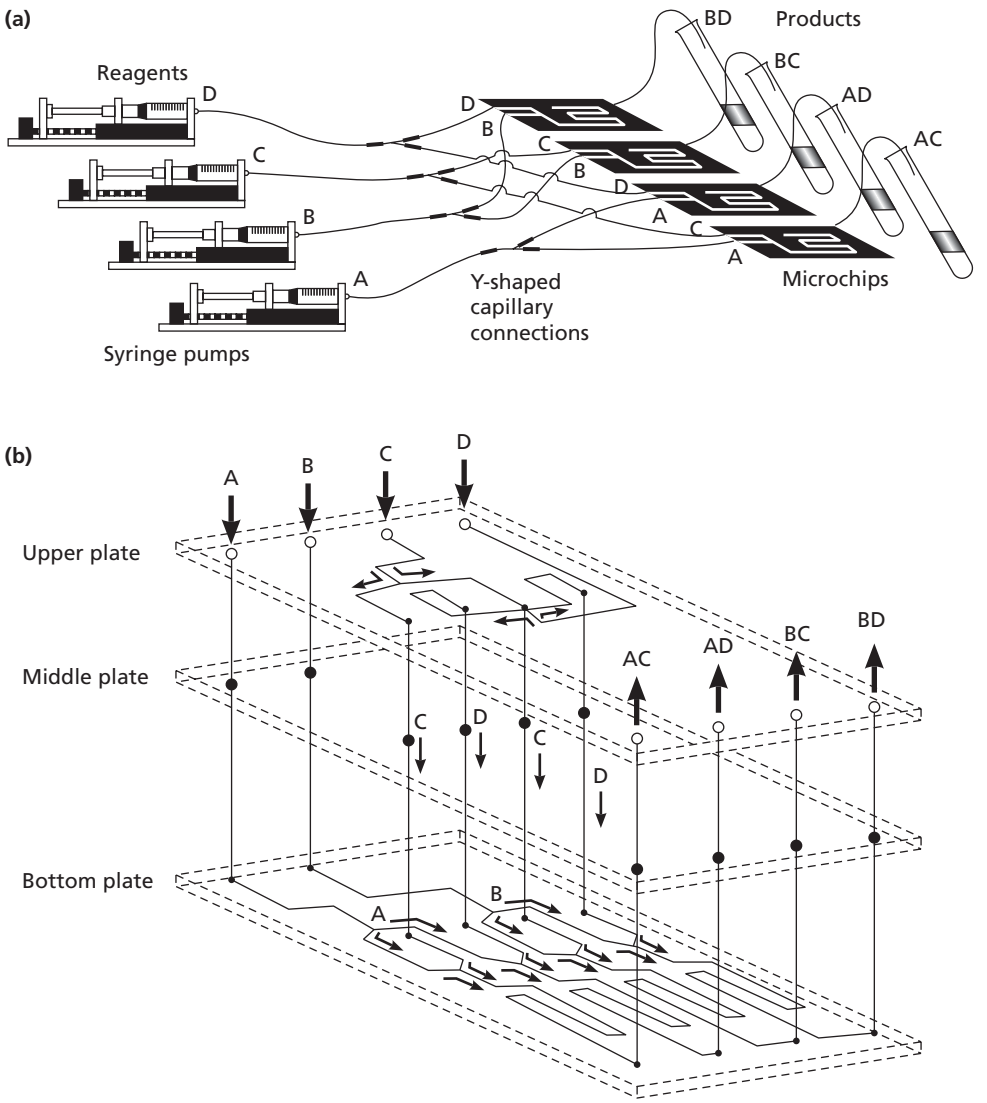


FIGURE 16.27 Parallel synthesis on a microchip. Parallel synthesis of four products using (**a**) four separate two-dimensional microchips and (**b**) a three-dimensional microchip. Reprinted by permission from Macmillan Publishers Ltd: *Nature Reviews Drug Discovery* (5, 210–18) 2006.

KEY POINTS

- In parallel synthesis, a reaction or series of reactions is carried out in a series of wells to produce a range of analogues. Each reaction well contains a single product.
- Parallel synthesis can be carried out on solid phase or in solution.
- Parallel synthesis allows the synthesis of a large number of easily identifiable analogues which can be tested quickly and easily, speeding up the optimization process.
- Solid phase extraction is often used in parallel synthesis for work-up. It involves the use of columns to remove impurities and excess reagents.
- An aqueous phase can be separated from an organic phase using phase separation columns or by freezing the aqueous phase onto a solid surface.
- Catch and release strategies involve reagents which are linked to a solid support. Reactants are taken out of solution when they react with the reagent and are then released when a subsequent reaction takes place.
- Solid supported reagents are easily removed at the end of a reaction. The potential toxicity of the reagent or its by-product is reduced when attached to a solid support.
- Microwave technology can prove advantageous over conventional heating.

16.6 Combinatorial synthesis

In combinatorial synthesis, mixtures of compounds are deliberately produced in each reaction vessel, allowing chemists to produce thousands, and even millions, of novel structures in the time that they would take to synthesize a few dozen by conventional means. This method of synthesis goes against the grain of conventional organic synthesis where chemists set out to produce a single identifiable structure which can be purified and characterized. The structures in each reaction vessel of a combinatorial synthesis are not separated and purified, but are tested for biological activity as a whole. If there is no activity, then there is no need to continue studies on that mixture and it is stored. If activity *is* observed, then one or more components in the mixture are active, although false-positives can sometimes be an issue (section 12.3.5). Overall, there is an economy of effort, as a negative result for a mixture of 100 compounds saves the effort of synthesizing, purifying, and identifying each component of that mixture. However, identifying the active component of an active mixture is not straightforward.

In a sense, combinatorial synthesis can be looked upon as the synthetic equivalent of nature's chemical pool. Through evolution, nature has produced a huge number and variety of chemical structures, some of which are biologically active. Traditional medicinal chemistry dips into that pool to pick out the **active principles** and develop them. Combinatorial synthesis produces pools of purely synthetic structures that we can dip into for active compounds. The diversity of structures from the natural pool is far greater than that likely to be achieved by combinatorial synthesis, but isolating, purifying, and identifying new agents from natural sources is a relatively slow process and there is no guarantee that a lead compound will be discovered against a specific drug target. The advantage of combinatorial chemistry is the fact that it produces new compounds faster than those derived from natural sources and can produce a diversity not found in the traditional banks of synthetic compounds held by pharmaceutical companies.

A few words of caution should be made here with regard to negative assays. There is always the possibility that a combinatorial mixture does not contain all the structures expected. This can happen if some of the starting materials or intermediates in the synthesis do not react as expected. A negative assay would then lead to the conclusion that these compounds are inactive when they are not actually present. This could mean that an active compound is missed. Assays might also be affected adversely if the individual components of a mixture interact with each other or have conflicting activities.

16.6.1 The mix and split method in combinatorial synthesis

A combinatorial synthesis is designed to produce a mixture of products in each reaction vessel, starting with a wide range of starting materials and reagents. This does not mean that all possible starting materials are thrown together in one reaction flask. If this was done, a black tarry mess would result. Instead, molecular structures are synthesized on solid supports, such as beads. Each individual bead may contain a large number of such molecules, but all the molecules on that bead are identical—'the one-bead-one-compound concept'. Different beads have different structures attached and can be mixed together in a single vial such that the molecules attached to the beads undergo the same reaction. In this way, each vial contains a mixture of structures, but each structure is physically distinct from the others as it is attached to a different bead.

Planning has to go into designing a combinatorial synthesis to minimize the effort involved and to maximize the number of different structures obtained. The strategy of **mix and split** is a crucial part of this. As an example, suppose we wish to synthesize all the possible dipeptides of five different amino acids. Using orthodox chemistry, we would synthesize these one at a time. There are 25 possible dipeptides and so we would have to carry out 25 separate experiments (Fig. 16.28).

Using a mix and split strategy the same products can be obtained with far less effort (Fig. 16.29). First of all, the beads are split between five reaction vials. The first amino acid is attached to the beads using a different amino acid for each vial. The beads from all five flasks are collected, mixed together, then split back into the five vials. This means that each vial now has the same mixture. The

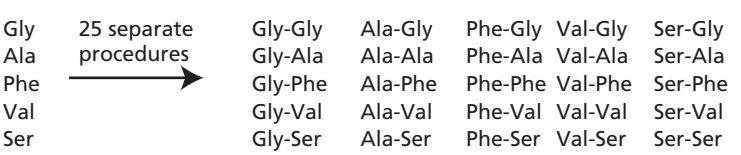


FIGURE 16.28 The possible dipeptides that can be synthesized from five different amino acids. Each procedure involves protection, coupling, and de-protection stages.

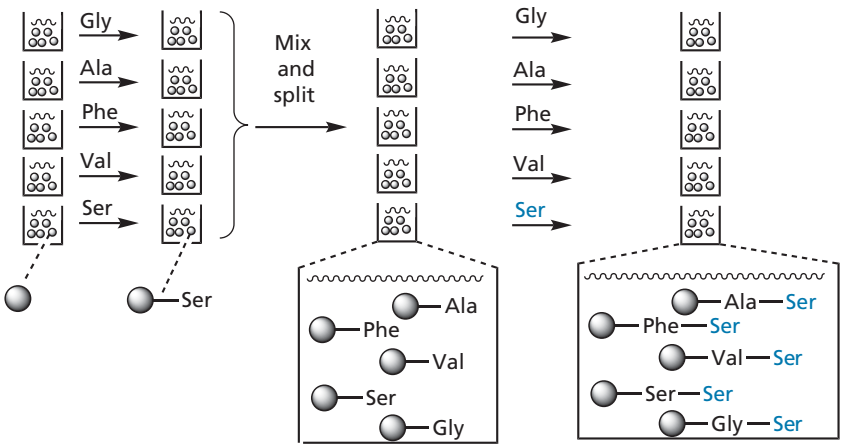


FIGURE 16.29 Synthesis of five different dipeptides using the mix and split strategy. Note that the addition of each amino acid involves protection, coupling, and de-protection steps.

second amino acid is now coupled with a different amino acid being used for each vial. Each vial now contains five different dipeptides with no one vial containing the same dipeptide. Each of the five mixtures can now be tested for activity. If the results are positive, the emphasis is on identifying which of the dipeptides is active. If there is no activity present, the mixture can be ignored.

In studies such as these, one can generate large numbers of mixtures, many of which are inactive. However, these mixtures are not discarded. Although they may not contain a lead compound on this particular occasion, they may provide the necessary lead compound for a different target in medicinal chemistry. Therefore, all the mixtures (both active and inactive) resulting from a combinatorial synthesis are stored as **compound libraries**. The example above produced 25 compounds in 5 mixtures. However, combinatorial synthesis can be used to produce several thousands of compounds.

16.6.2 Structure determination of the active compound(s)

The direct structural determination of components in a compound mixture is no easy task, but advances have been made in obtaining interpretable mass, nuclear magnetic resonance (NMR), Raman, infrared, and ultraviolet spectra on products attached to a single resin bead. Peptides can be sequenced while still attached to the bead. Each 100-μm bead contains about 100 pmole of peptide, which is enough for microsequencing. With non-peptides, the structural determination of an active compound can be achieved by **recursive deconvolution** methods. Alternatively, **tagging** procedures can be used during the synthesis.

For additional material see Web article 6: *deconvolution.*

16.6.2.1 Tagging

In this process, two molecules are built up on the same bead. One of these is the intended structure, the other is a molecular tag (usually a peptide or oligonucleotide) which will act as a code for each step of the synthesis. For this to work, the bead must have a multiple linker capable of linking both the target structure and the molecular tag. A starting material is added to one part of the linker, and an encoding amino acid or nucleotide to another part. After each subsequent stage of the combinatorial synthesis, an amino acid or nucleotide is added to the growing tag to indicate what reagent was used. One example of a multiple linker is called the **safety catch acid-labile linker** (**SCAL**) (Fig. 16.30), which includes lysine and tryptophan. Both these amino acids have a free amino group.

The target structure is constructed on the amino group of the tryptophan moiety and, after each stage of the synthesis, a tagging amino acid is built on to the amino groups of the lysine moiety. Figure 16.31 illustrates the procedure for a synthesis involving three reagents, so that by the end of the process there is a tripeptide tag where each amino acid defines the identity of the variable groups R, R′, and R″ in the target structure.

The non-peptide target structure can be cleaved by reducing the two sulphoxide groups in the safety catch linker, then treating with acid. Under these conditions, the tripeptide sequence remains attached to the bead and can be sequenced on the bead to identify the structure of the compound that was released.

The same strategy can be used with an oligonucleotide as the tagging molecule. The oligonucleotide can

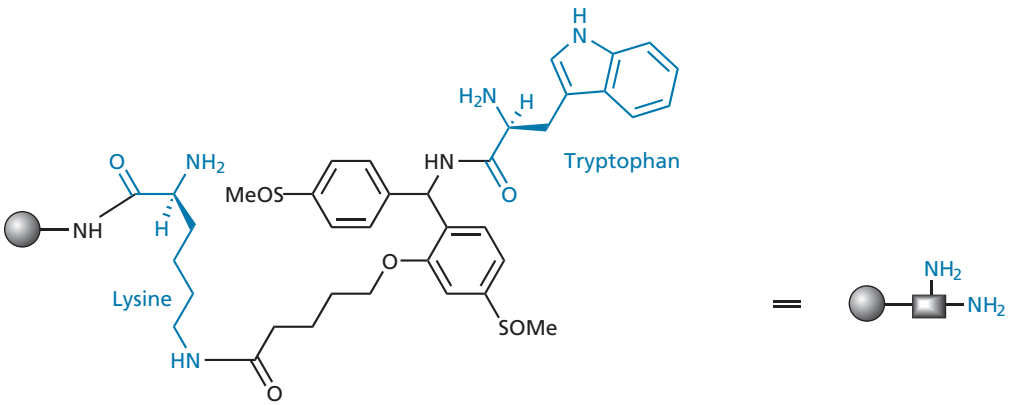


FIGURE 16.30 Safety catch acid-labile linker (SCAL).

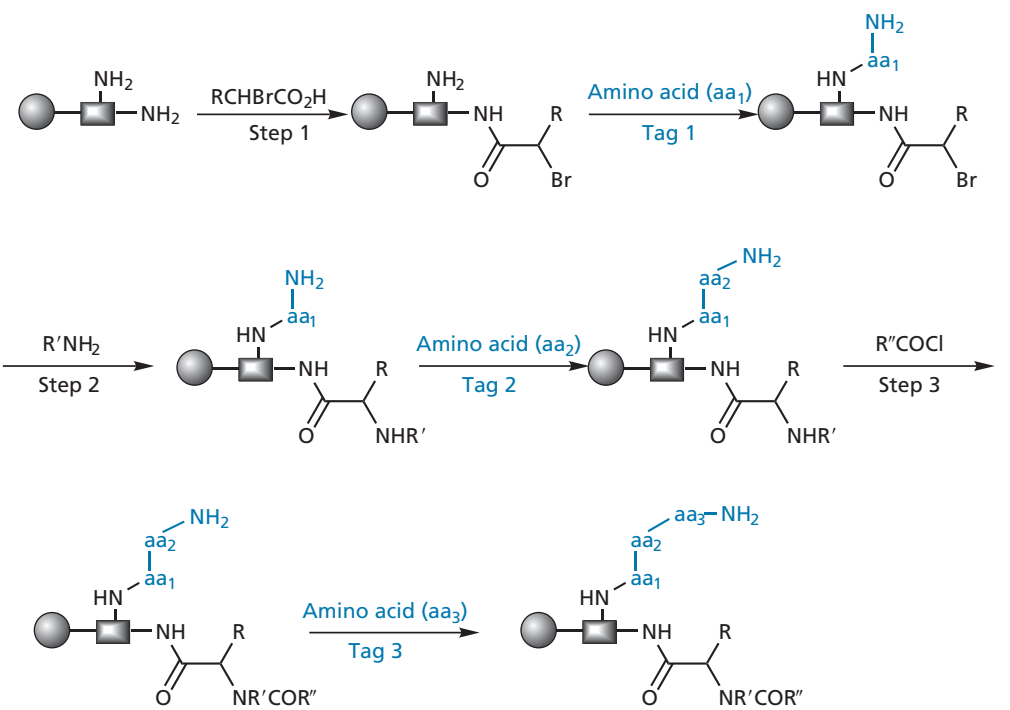


FIGURE 16.31 Tagging a bead to identify the structure being synthesized. Note that the reaction sequence has been simplified here to illustrate the principle of tagging. Amino acids are *N*-protected when coupled and the protecting group is removed before the next coupling. Coupling agents are also present. An orthogonal protection strategy is also required to distinguish between the amino groups of the safety catch acid-labile linker (SCAL).

be amplified by replication and the code read by DNA sequencing.

There are drawbacks to tagging processes. They are time-consuming and require elaborate instrumentation. Building the coding structure itself also adds extra restraints on the protection strategies that can be employed and may impose limitations on the reactions that can be used. In the case of oligonucleotides, their inherent instability can prove a problem. Another possible problem with tagging is the possibility of an unexpected reaction taking place, resulting in a different structure from that expected. Nevertheless, the tagging procedure is still valid as it identifies the starting materials and the reaction conditions, and when these are repeated on larger scale any unusual reactions would be discovered.

These tagging methods require the use of a specific molecular tag to represent each reagent used in the synthesis. Moreover, the resultant molecular tag has to be sequenced at the end of the synthesis. A more efficient method of tagging and identifying the final product is to use some form of encryption or **bar code**. For example, it is possible to identify which one of seven possible reagents has been used in the first stage of a synthesis with the use of only three molecular labels (A–C). This is achieved by adding different combinations of the three tags to set up a triplet code on the bead. Thus, adding just one of the tags (A, B, or C) will allow the identification of three of the reagents. Adding two of the tags at the same time allows the identification of another three reagents and adding all three tags at the same time allows the identification of a seventh reagent. The presence (1) or absence (0) of the tag forms a triplet code; the presence of a single molecular tag (A, B, or C) gives the triplet codes (100, 010, and 001); the presence of two different tags is indicated by another three triplet codes (110, 101, 011); and the presence of all three tags is represented by 111. The tags are linked to the bead by means of a photocleavable bond, so irradiating the bead releases all the tags. These can then be passed through a gas chromatograph and identified by their retention time.

Three different molecular tags could now be used to represent seven reagents in the second stage and so on. All the tags used to represent the second reagent would have longer retention times than the tags used to represent the first reagent. Similarly, all subsequent tags would have longer retention times. Once the synthesis is complete, all the tags are released simultaneously and passed through the gas chromatograph as before. The 'bar code' is then read from the chromatograph in one go, not only identifying the reagents used, but the order in which they were used (Fig. 16.32).

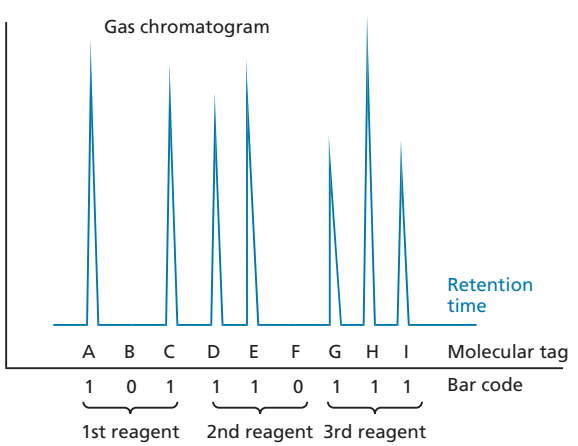


FIGURE 16.32 Identification of reagents and order of use by bar-coding.

16.6.2.2 Photolithography

Photolithography is a technique that permits miniaturization and spatial resolution such that specific products are synthesized on a plate of immobilized solid support. In the synthesis of peptides, the solid support surface contains an amino group protected by the photolabile **nitroveratryloxycarbonyl** (**NVOC**)-protecting group (Fig. 16.33). Using a mask, part of the surface is exposed to light resulting in de-protection of the exposed region. The plate is then treated with a protected amino acid and the coupling reaction takes place only on the region of the plate which has been de-protected. The plate is then washed to remove excess amino acid. The process can be repeated on a different region using a different mask, and so different peptide chains can be built on different parts of the plate; the sequences are known from the record of masks used.

Incubation of the plate with a protein receptor can then be carried out to detect active compounds which bind to the binding site of the receptor. A convenient method to assess such interactions is to incubate the plate with a fluorescently tagged receptor. Only those regions of the plate which contain active compounds will bind to the receptor and fluoresce. The fluorescence intensity can be measured using fluorescence microscopy and is a measure of the affinity of the compound for the receptor. Alternatively, testing can be carried out such that active compounds are detected by radioactivity or chemiluminescence.

The photo de-protection described earlier can be achieved in high resolution. At a 20-μm resolution, plates can be prepared with 250,000 separate compounds per square centimetre.

16.6.3 Dynamic combinatorial synthesis

Dynamic combinatorial synthesis is an exciting development which has been used as an alternative to the classic mix and split combinatorial syntheses in the search for new lead compounds. The aim of dynamic combinatorial synthesis is to synthesize all the different compounds in one flask at the same time, screen them *in situ* as they are being formed, and thus identify the most active compound in a much faster time period (Box 16.1). How can this be achieved? There are several important principles which are followed.

- The best way of screening the compounds is to have the desired target present in the reaction flask along with the building blocks. This means that any active compounds can bind to the target as soon as they are formed. The trick is then to identify which of the products are binding.
- The reactions involved should be reversible. If this is the case, a huge variety of products are constantly being formed in the flask then breaking back down

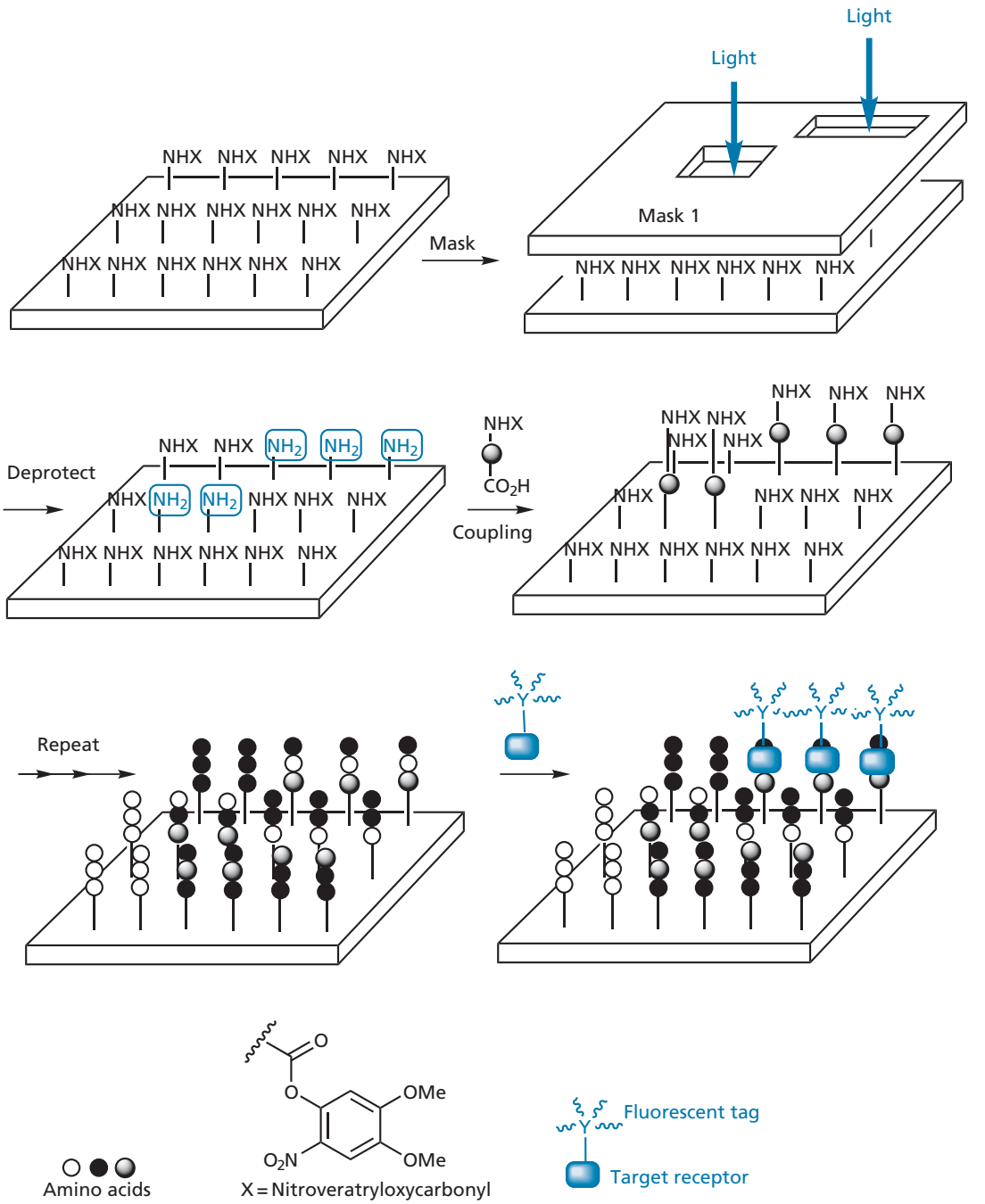


FIGURE 16.33 Photolithography.

into their constituent building blocks. The advantage of this may not seem obvious, but it allows the possibility of 'amplification', where the active compound is present to a greater extent than the other possible products. By having the target present, active compounds become bound and are removed effectively from the equilibrium mixture. The equilibrium is now disturbed such that more of the active product is formed. Thus, the target serves not only to screen for active compounds, but also to amplify them.

- In order to identify the active compounds, it is necessary to 'freeze' the equilibrium reaction such that it no longer takes place. This can be done by carrying out a further reaction which converts all the equilibrium products into stable compounds that cannot revert back to starting materials.

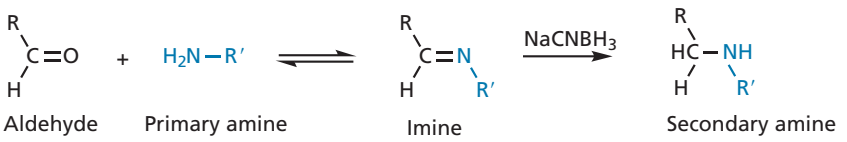


FIGURE 16.34 Example of dynamic combinatorial synthesis.

A simple example of dynamic combinatorial synthesis involved the reversible formation of imines from aldehydes and primary amines (Fig. 16.34). A total of three aldehydes and four amines were used in the study (Fig. 16.35), allowing the possibility of 12 different imines in the equilibrium mixture.

The building blocks were mixed together with the target enzyme **carbonic anhydrase** and allowed to interact. After a suitable period of time, sodium cyanoborohydride was added to reduce all the imines present to secondary amines so that they could be identified (Fig. 16.34). The mixture was separated by reverse-phase high-performance liquid chromatography (HPLC), allowing each product to be quantified and identified. These results were compared with those obtained when the experiment was carried out in the absence of carbonic anhydrase, making it possible to identify which products had been amplified. In this experiment, the sulphonamide shown in Fig. 16.36 was significantly amplified, which demonstrated that the corresponding imine was an active compound.

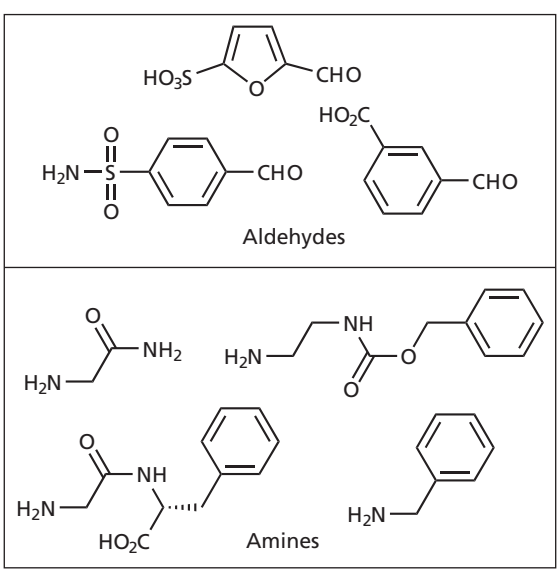


FIGURE 16.35 Aldehyde and amine building blocks used in the dynamic combinatorial synthesis of imines.

The above example illustrates a simple case involving one reaction and two sets of building blocks, but it is feasible to have more complex situations. For example, a molecule with two or more functional groups could be present to act as a scaffold on to which various substituents could be added from the building blocks available (Fig. 16.37). The use of a central scaffold has another benefit: it helps the amplification process. If the number of scaffold molecules present is equal to the number of target molecules, then the number of products formed cannot be greater than the number of targets available. If any of these products binds to the target, the effect on the equilibrium will be greater than if there were more products than targets available.

There are certain limitations to **dynamic combinatorial chemistry**:

- conditions must be chosen such that the target does not react chemically with any of the building blocks or is unstable under the reaction conditions used;
- the target is normally in an aqueous environment, so the reactions have to be carried out in aqueous solution;
- the reactions themselves have to undergo fast equilibration rates to allow the possibility of amplification;
- it is important to avoid using some building blocks that are more likely to react than others, as this would bias the equilibrium toward particular products and confuse the identification of the amplified product.

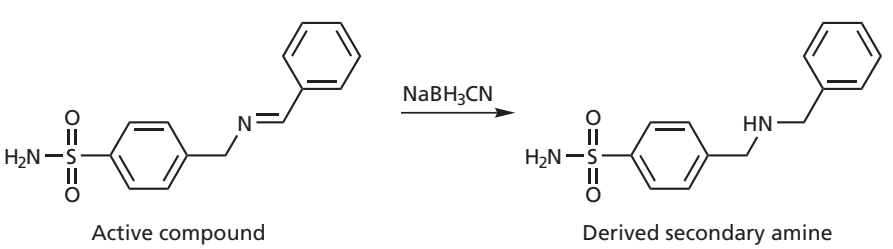


FIGURE 16.36 Amplified imine and the amine obtained from reduction.

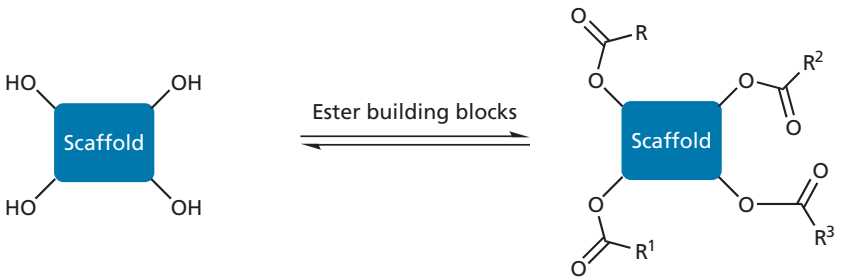


FIGURE 16.37 Use of a scaffold molecule.

BOX 16.2 Dynamic combinatorial synthesis of vancomycin dimers

Vancomycin is an antibiotic that works because it masks the building blocks required for bacterial cell wall synthesis (section 19.5.5.2). Binding takes place specifically between the antibiotic and a peptide sequence (L-Lys–D-Ala–D-Ala) which is present in the building block. It is also known that this binding promotes dimerization of the vancomycin–target complex, which suggests that covalently linked vancomycin dimers might be more effective antibacterial agents than vancomycin itself. A dynamic combinatorial synthesis was carried out to synthesize a variety of different vancomycin dimers covalently linked by bridges of different lengths. The vancomycin monomers used had been modified such that they contained long-chain alkyl substituents with double bonds at the end. Reaction between the double bonds in the presence of a catalyst then led to bridge formation through a reaction known as olefin metathesis (Fig. 1).

The tripeptide target was present to accelerate the rate of bridge formation and to promote formation of vancomycin dimers having the ideal bridge length. As shown in Fig. 2, the vancomycin monomers bind the tripeptide, which encourages the self-assembly of non-covalently-linked dimers. Once formed, those dimers having the correct length of substituent are more likely to react together to form the covalent bridge (Fig. 2).

Having established the optimum length of bridge, another experiment was carried out on eight vancomycin monomers which had the correct length of 'tether' but varied slightly in their structure. The mixture of 36 possible products was analysed by mass spectrometry to indicate the relative proportion of each dimer formed. Eleven of the 36 compounds were then synthesized separately and it was found that their antibacterial activity matched their level of amplification, i.e. the compounds present in greater quantities had the greater activity.

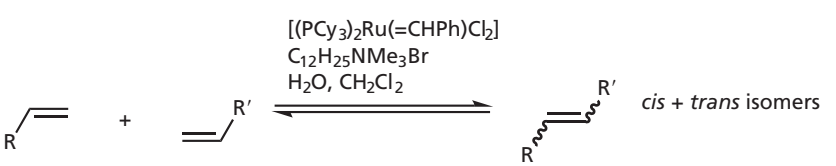


FIGURE 1 Olefin metathesis.

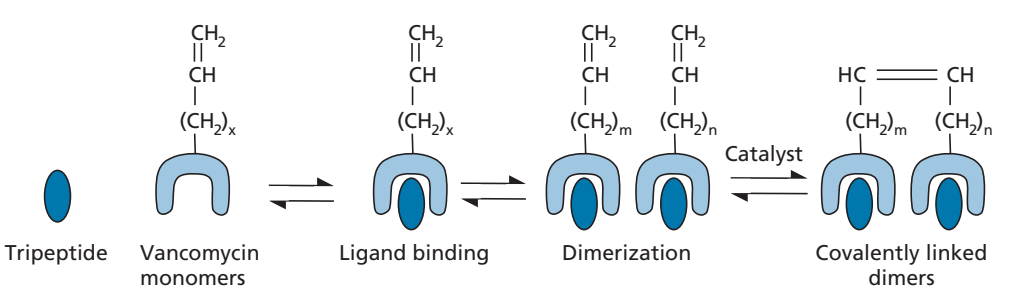


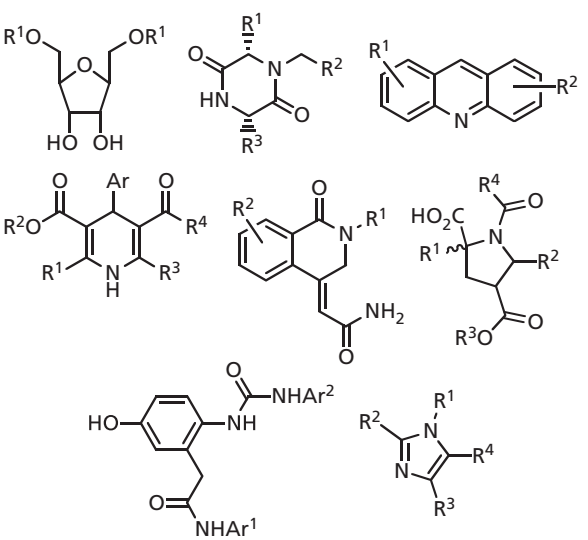
FIGURE 2 Formation of covalently linked dimers.

KEY POINTS

- Most combinatorial syntheses are carried out using automated or semi-automated synthesizers.
- The mix and split method allows the efficient synthesis of large numbers of compounds with a minimum number of operations.
- The compounds synthesized in a combinatorial synthesis are stored as compound libraries.
- Tagging involves the construction of a tagging molecule on the same solid support as the target molecule. Tagging molecules are normally peptides or oligonucleotides. After each stage of the target synthesis, the peptide or oligonucleotide is extended and the amino acid or nucleotide used defines the reactant or reagent used in that stage.
- Photolithography is a technique involving a solid support surface containing functional groups protected by photolabile groups. Masks are used to reveal defined areas of the plate to light, thus removing the protecting groups and allowing a reactant to be linked to the solid support. A record of the masks used determines what reactions have been carried out at different regions of the plate.
- Combinatorial synthesis has been used for the synthesis of peptides, peptoids, and heterocyclic structures. Most organic reactions are feasible.
- Dynamic combinatorial chemistry involves the equilibrium formation of a mixture of compounds in the presence of a target. Binding of a product with the target amplifies that product in the equilibrium mixture.

QUESTIONS

1. Identify three stages of the drug discovery, design, and development process where combinatorial chemistry or parallel synthesis is of importance.
2. A pharmaceutical laboratory wishes to synthesize all the possible dipeptides containing the amino acids tyrosine, lysine, phenylalanine, and leucine. Identify the number of possible dipeptides and explain how the laboratory would carry this out using combinatorial techniques.
3. What particular precautions have to be taken with the amino acids tyrosine and lysine in the above synthesis?
4. Identify the advantages and disadvantages of the following structures as scaffolds.



5. You wish to carry out the combinatorial synthesis shown in Fig. 16.31 using bar-coding techniques rather than the conventional tagging scheme shown in the figure. You have nine molecules suitable for tagging purposes (A–I), seven bromo acids (B1–B7), seven amines (A1–A7), and seven acid chlorides (C1–C7). Construct a suitable coding system for the synthesis.
6. Based on your coding scheme from Question 5, what product is present on the bead if the released tags resulted in the gas chromatogram shown in Fig. 16.32.

FURTHER READING

Beck-Sickinger, A. and Weber, P. (2002) *Combinatorial Strategies in Biology and Chemistry*. John Wiley and Sons, New York.

Bhalay, G., Dunstan, A., and Glen, A. (2000) Supported reagents: Opportunities and limitations. *Synlett* **12**, 1846–1859.

Booth, R.J., and Hodges, J.C. (1997) Solid-supported quenching reagents for parallel purification. *Journal of the American Chemical Society* **119**, 4882–4886.

Braeckmans, K., De Smedt, S. C., Leblans, M., Pauwels, R., and Demeester, J. (2002) Encoding microcarriers: present and future technologies. *Nature Reviews Drug Discovery* **1**, 447–456.

DeWitt, S. H., Kiely, J. S., Stankovic, C. J., Schroeder, M. C., Cody, D. M., and Pavia, M. R. (1993) 'Diversomers': An approach to nonpeptide, nonoligomeric chemical diversity. *Proceedings of the National Academy of Sciences of the USA* **90**, 6909–6913.

Dittrich, P. S. and Manz, A. (2006) Lab-on-a-chip: microfluidics in drug discovery. *Nature Reviews Drug Discovery* **5**, 210–218.

Dobson, C. M. (2004) Chemical space and biology. *Nature,* **432**, 824–828.

Dolle, R. E. (2003) Comprehensive survey of combinatorial library synthesis: 2002. *Journal of Combinatorial Chemistry* **5**, 693–753.

Geysen, H. M., Schoenen, F, Wagner, D., and Wagner, R. (2003) Combinatorial compound libraries for drug discovery: an ongoing challenge. *Nature Reviews Drug Discovery* **2**, 222–230.

Guillier, F., Orain, D., and Bradley, M. (2000) Linkers and cleavage strategies in solid-phase organic synthesis and combinatorial chemistry. *Chemical Reviews* **100**, 2091–2158.

Houlton, S. (2002) Sweet synthesis. *Chemistry in Britain* April, 46–49.

Jung, G. (ed.) (1996) *Combinatorial Peptide and Nonpeptide Libraries*. VCH, Weinheim.

Kappe, C. O. (2004) Controlled microwave heating in modern organic synthesis. *Angewandte Chemie International Edition* **43**, 6250–6284..

Le, G. T., Abbenante, G., Becker, B., Grathwohl, M., Halliday, J., Tometzki, G, et al. (2003) Molecular diversity through sugar scaffolds. *Drug Discovery Today* **8**, 701–709.

Ley, S. V. and Baxendale, I. R. (2002) New tools and concepts for modern organic synthesis. *Nature Reviews Drug Discovery* **1**, 573–586.

Mavandadi, F. and Pilotti, Å. (2006) The impact of microwave-assisted organic synthesis in drug discovery. *Drug Discovery Today* **11**, 165–174.

Nicolaou, K. C., Hughes, R., Cho, S. Y., Winssinger, N., Smethurst, C., Labischinski, H., and Endermann, R. (2000) Target-accelerated combinatorial synthesis and discovery of highly potent antibiotics effective against vancomycin-resistant bacteria. *Angewandte Chemie, International Edition* **39**, 3823–3828.

Ramstrom, O. and Leh, J.-M. (2002) Drug discovery by dynamic combinatorial libraries. *Nature Reviews Drug Discovery* **1**, 26–36.

Reader, J. C. (2004) Automation in medicinal chemistry. *Current Topics in Medicinal Chemistry* **4**, 671–686.

Terret, N. K. (1998) *Combinatorial Chemistry*. Oxford University Press, Oxford.

Titles for general further reading are listed on p. 763. See also articles in Journal of Combinatorial Chemistry.

For additional material see Web article 20: modern chemistry techniques in medicinal chemistry

For additional material see Web article 21: microwave technology applied to medicinal chemistry

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 specialized textbooks and journal articles (see Further reading).

17.1 Molecular and quantum mechanics

The various operations 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. 17.16).

These calculations require data or parameters that 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 as 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. However, it is expensive on computer time 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.

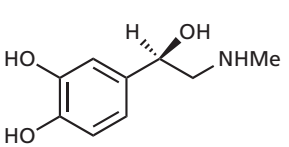
17.2 Drawing chemical structures

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 **IsisDraw**, are available which 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 nuclear magnetic resonance (NMR) chemical shifts, melting point, freezing point, log *P* value, molar refractivity, and heat of formation (Fig. 17.1).

17.3 Three-dimensional 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 **Chem3D**, **Alchemy**, **Sybyl**, **Hyperchem**, **Discovery Studio Pro**, **Spartan**, and **CAChe**. The 3D model can be made by constructing the molecule atom

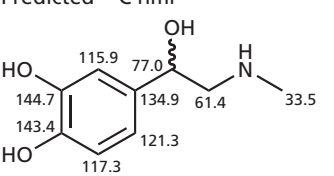


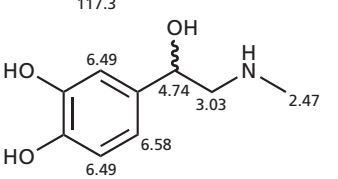
1-(3,4-Dihydroxyphenyl)-2-methylaminoethanol

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.

by atom, and bond by bond. It is also possible to automatically convert a two-dimensional (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 pasted into Chem3D, 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. 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).

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).

Test your understanding and practise your molecular modelling with Exercise 17.1.

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.

Test your understanding and practise your molecular modelling with Exercise 17.2.

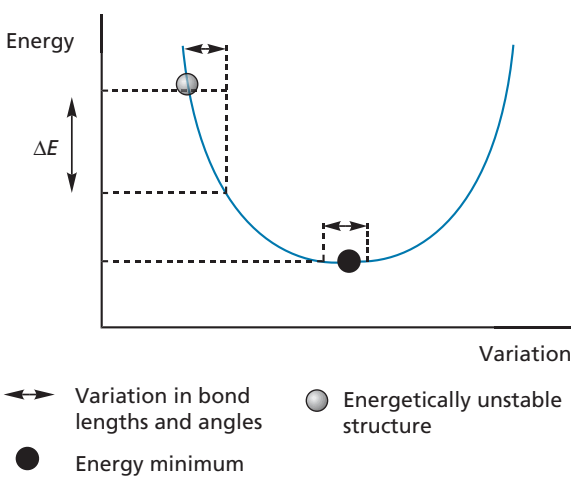


FIGURE 17.3 Energy minimization.

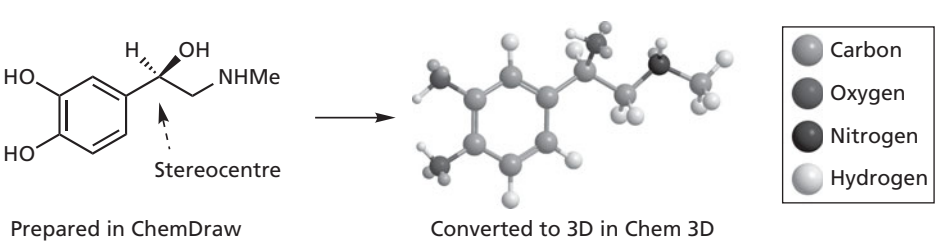
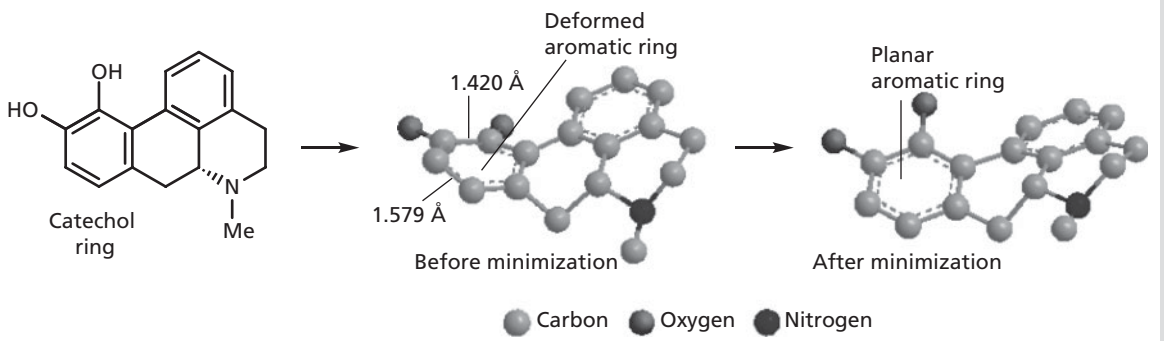


FIGURE 17.2 Conversion of a 2D drawing to a 3D model.

BOX 17.1 Energy minimizing apomorphine

A 2D structure of **apomorphine** was converted to a 3D structure using Chem3D. 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.



Energy minimization carried out on apomorphine.

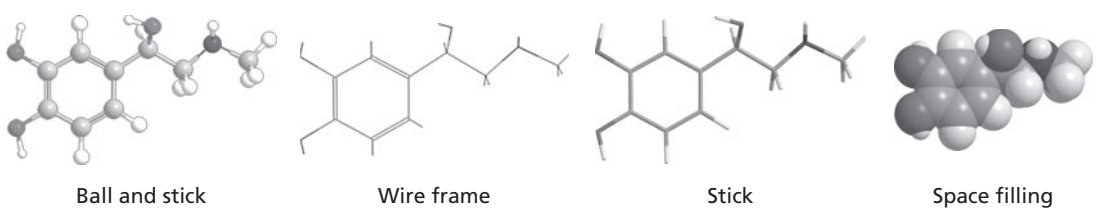


FIGURE 17.4 Different methods of visualizing molecules.

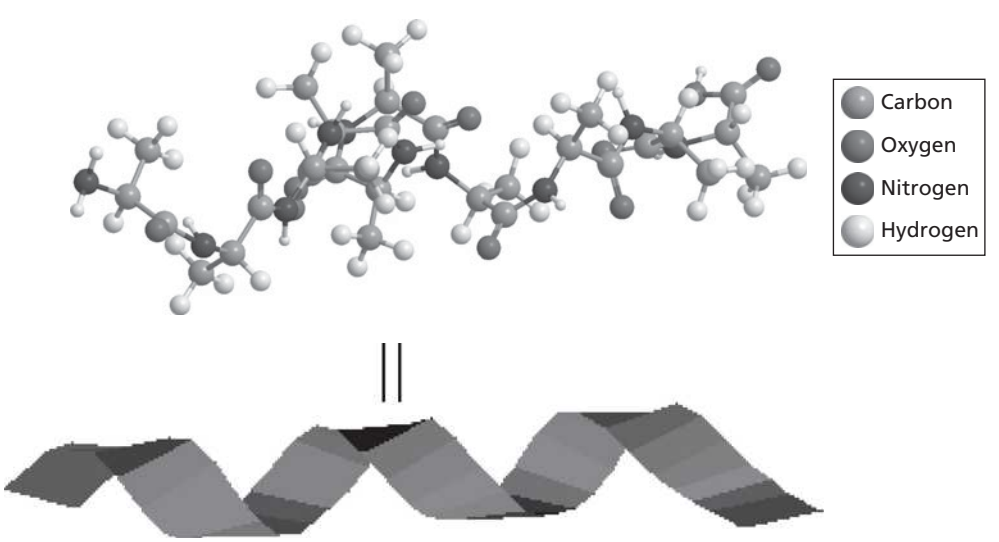


FIGURE 17.5 Ribbon representation of a helical decapeptide (Chem 3D).

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 (see molecular modelling Exercise 17.1).

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 which are 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 in the following sections.

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 ones, 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.

Test your understanding and practise your molecular modelling with Exercise 17.3.

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

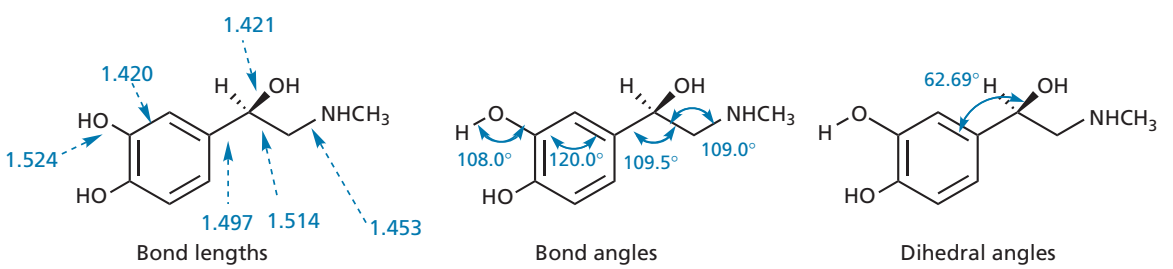


FIGURE 17.6 Molecular dimensions for adrenaline (Chem 3D).

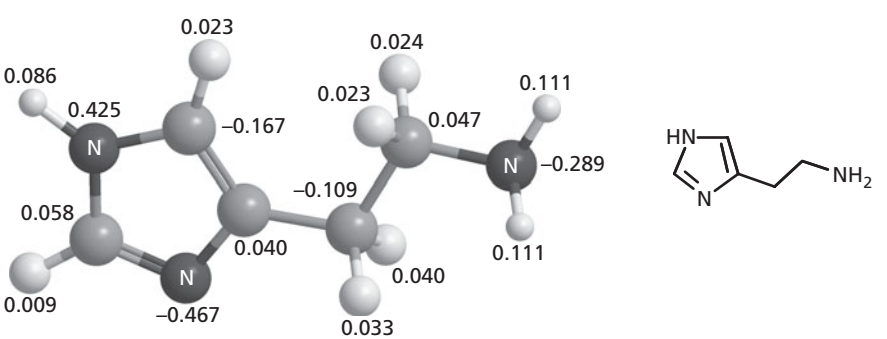


FIGURE 17.7 Partial charges for histamine.

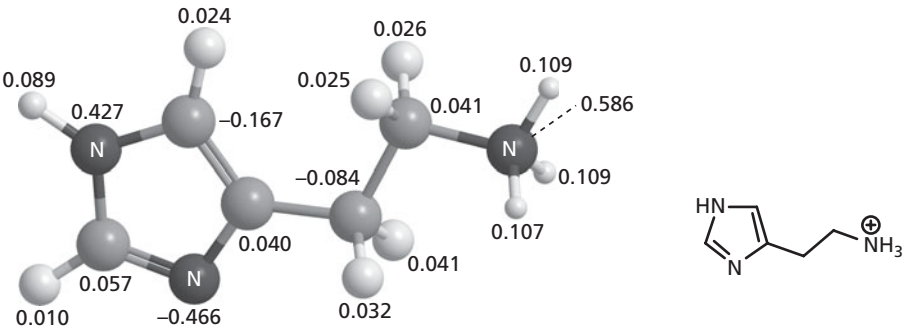


FIGURE 17.8 Charge distribution on the histamine ion.

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.

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, 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.

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. 17.9 was calculated using the semi-empirical method AM1. Another method of calculating MEPs is described in section 17.7.5.

An example of how electrostatic potentials have been used in drug design can be seen in the design of the **cromakalim** analogue (II; Fig. 17.10), where the cyanoaromatic ring was replaced by a pyridine ring. This was part of a study looking into analogues of cromakalim which 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

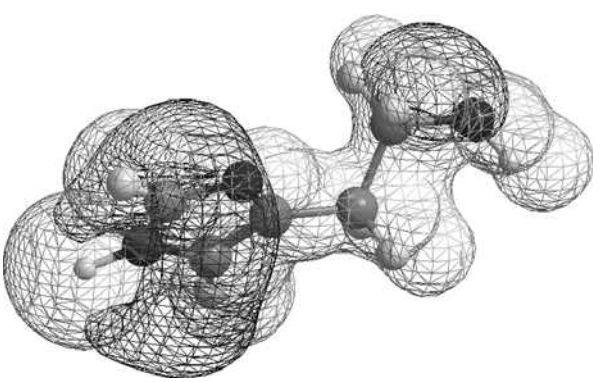
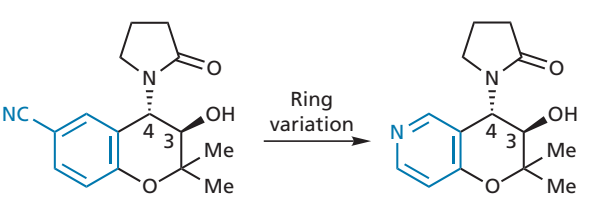


FIGURE 17.9 Molecular electrostatic potential for histamine.



Cromakalim II

FIGURE 17.10 Ring variation on cromakalim.

systems were calculated and compared with the parent bicyclic system (III; Fig. 17.11). 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. 17.12). The contours represent the various levels of the MEP and can be taken 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.

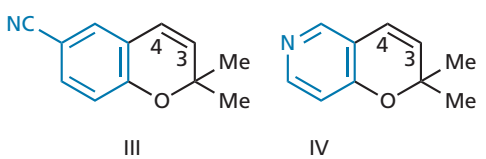


FIGURE 17.11 Bicyclic models in cromakalim study.

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. 17.13 (see also Box 17.2).

Test your understanding and practise your molecular modelling with Exercise 17.4.

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

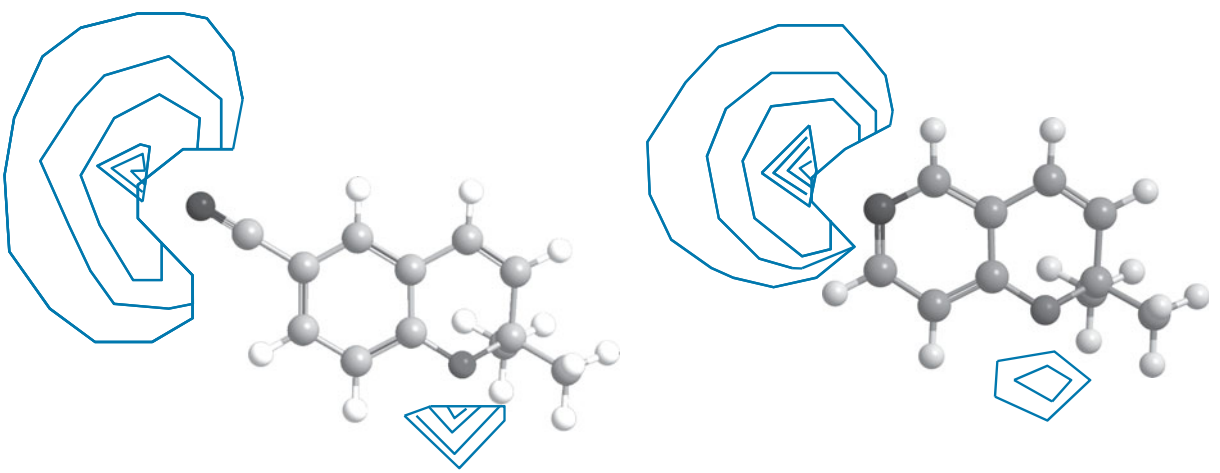


FIGURE 17.12 Molecular electrostatic potentials (MEPs) of bicyclic models (III) and (IV).

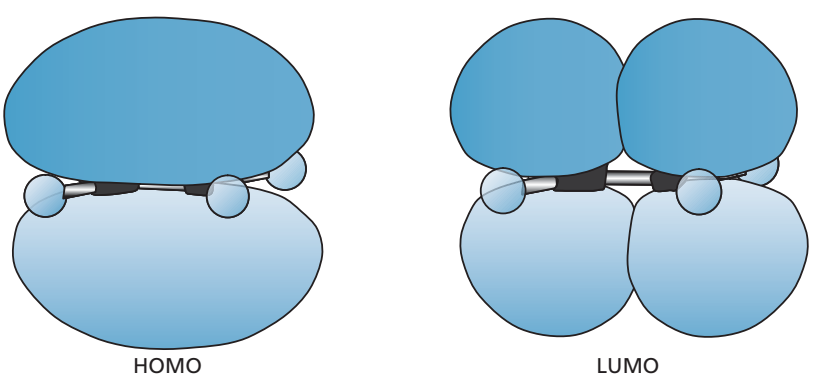


FIGURE 17.13 HOMO and LUMO molecular orbitals for ethene.

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.

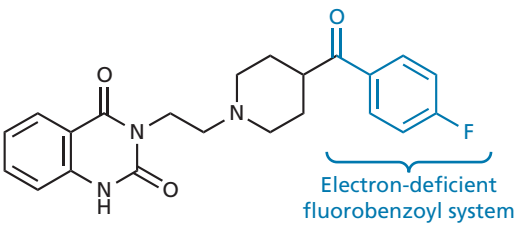


FIGURE 1 Ketanserin.

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.

Test your understanding and practise your molecular modelling with Exercise 17.5.

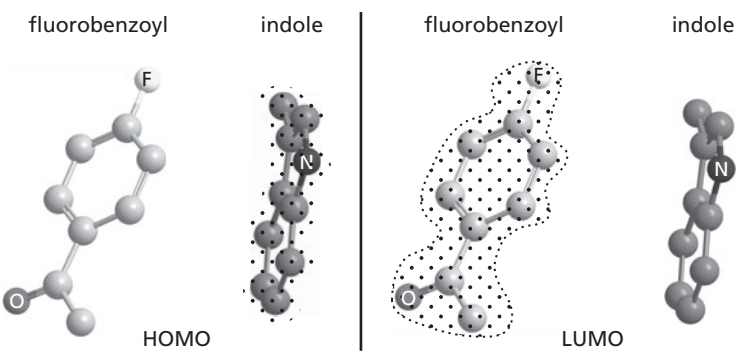


FIGURE 2 HOMO and LUMO molecular orbitals (dot surfaces) for the indole/fluorobenzoyl complex.

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 The use of grids in measuring molecular properties

Grids have become used extensively 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).

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 pre-constructed 3D lattice or grid (Fig. 17.14). 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

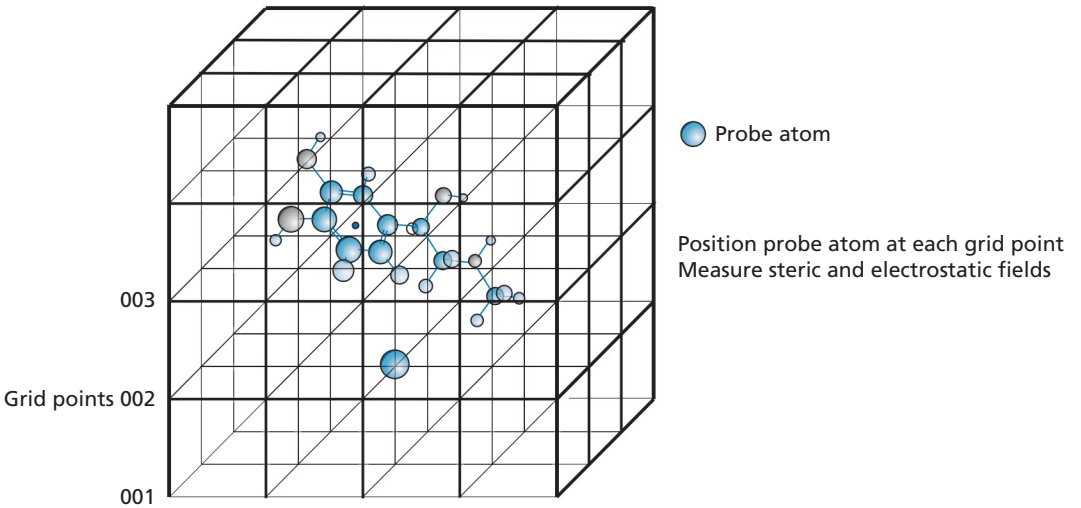


FIGURE 17.14 Measuring fields around a molecule by placing a probe atom at grid points.

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. Obviously, this 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 calculated quickly 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 and 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. However, the structure obtained is not necessarily the most stable conformation. This is because energy minimization stops as soon as it reaches the first stable conformation it finds, which will be the one closest to the starting structure. This is illustrated in Fig. 17.15, where the most stable conformation is separated from a less stable conformation by an energy saddle. If the 3D structure created initially 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.

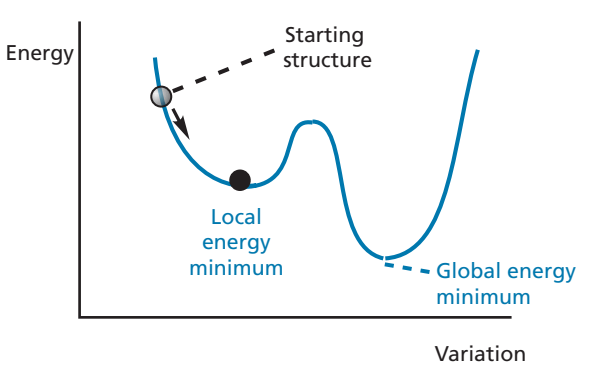


FIGURE 17.15 Local and global energy minima.

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 and potential energies would then be huge leading to atoms moving with excessive velocities and accelerations, and resulting in system failure.

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 distinct conformations that might be more stable than the initial conformation.

For example, the 2D drawing of **butane** shown in Fig. 17.16 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

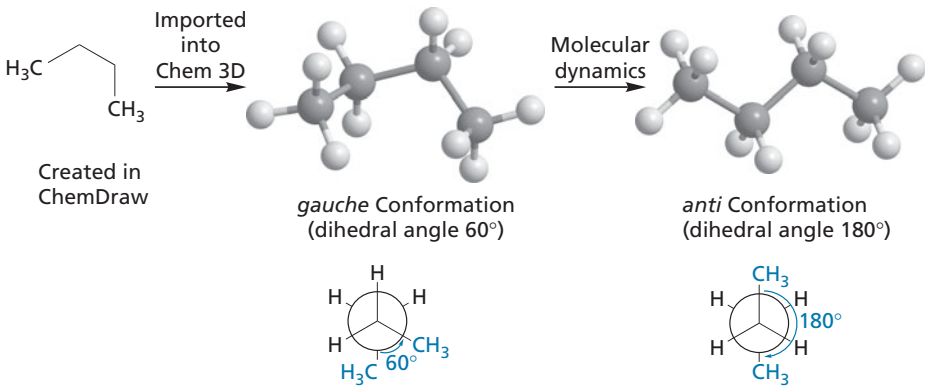


FIGURE 17.16 Use of molecular dynamics to find the most stable conformation.

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—see Box 17.3) or which would take too long to analyse by that process (large flexible molecules).

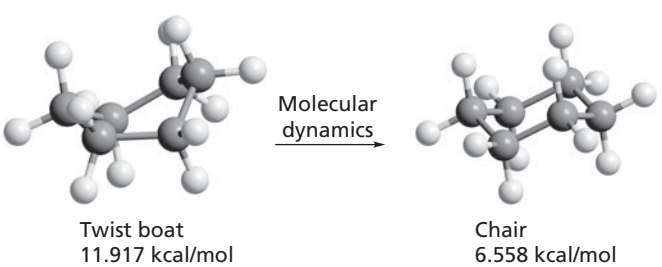
Finally, it has to be remembered that biomolecules in the real world are surrounded by water and that this can affect the relative stability of different conformations. Therefore, it is advisable to include water molecules in the modelling system before carrying out molecular dynamics experiments.

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 that 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. 17.17), revealing that the most stable conformation was the fully staggered one, whereas 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.

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 Chem3D.

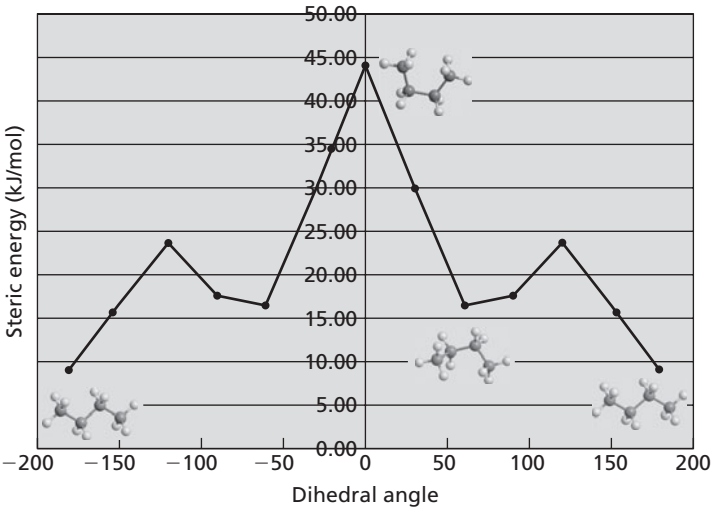


FIGURE 17.17 Graph showing relative stabilities of various butane conformations.

Test your understanding and practise your molecular modelling with Exercise 17.6.

Some modelling software packages can automatically identify all the rotatable single bonds in a structure. Bonds to hydrogen or simple substituents 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.

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 has 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 conformations which are all analysed separately. This means that the same amount of processing time is taken up by high- 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. 17.18) 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

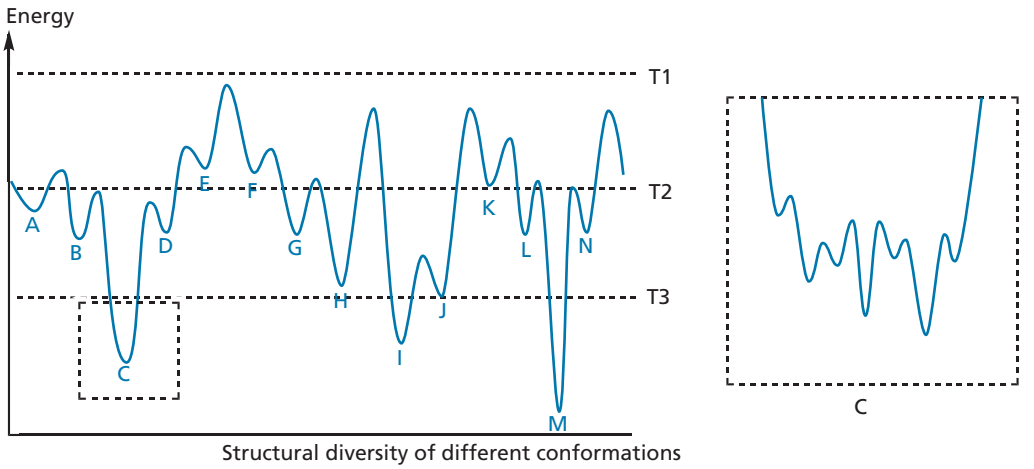


FIGURE 17.18 A 2D representation of conformational space versus structural diversity of conformations.

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 17.18. 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—owing 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 localized 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 as 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 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 of success.

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 as 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 area of conformational space containing the global minimum.

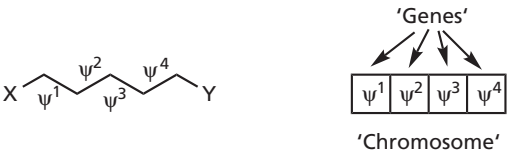


FIGURE 17.19 The representation of torsion angles as genes in chromosomes.

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 in 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 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. 17.19).

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 'crossover' or recombination process to generate new chromosomes where each chromosome has torsion angles contributed from each parent. In the example shown in Fig. 17.20, the crossover 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 crossovers, random 'single point' mutations are made on individual chromosomes. This corresponds to a random alteration of a single torsion angle within the chromosome. Both crossover 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.

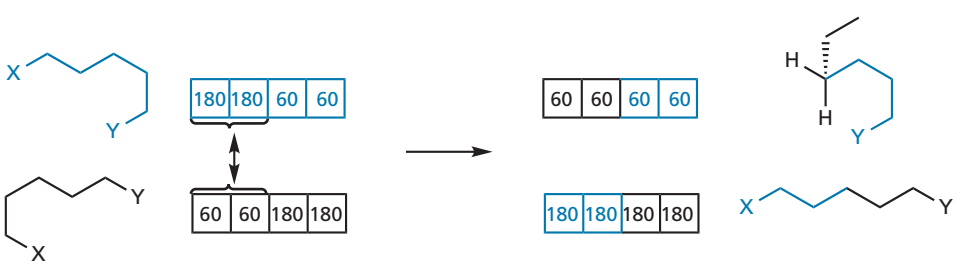


FIGURE 17.20 The crossover or recombination process to generate new conformations.

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

Evolutionary programming is slightly different from genetic algorithms in that crossovers 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 period of time, but owing to their random nature, they cannot be guaranteed to find the global conformation. 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 bonds—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 the alkaloid **cocaine** with the synthetic agent **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. 17.21, 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. 17.22. 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 are matched up, and finding 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.

Test your understanding and practise your molecular modelling with Exercise 17.7.

It is important to appreciate that the fitting process described earlier 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

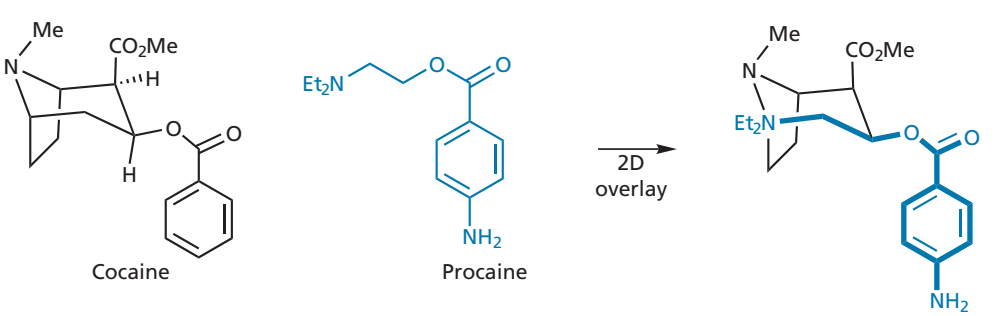


FIGURE 17.21 2D overlay of cocaine and procaine.

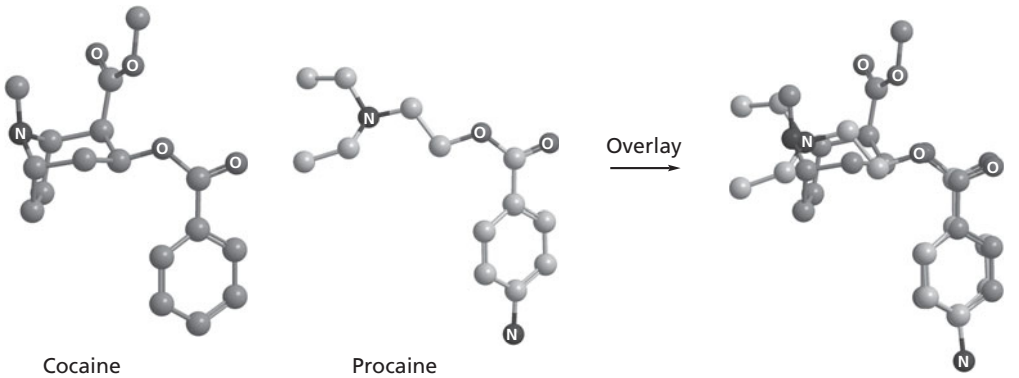


FIGURE 17.22 Overlay of cocaine and procaine using Chem3D.

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 centres that are normally involved in binding interactions (i.e. aromatic rings, hydrogen bond donors, hydrogen bond acceptors, positively charged 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. Of course, 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 or non-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 encountered frequently 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 stabilization 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 crystallized easily, 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 (see 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).

Test your understanding and practise your molecular modelling with Exercise 17.8.

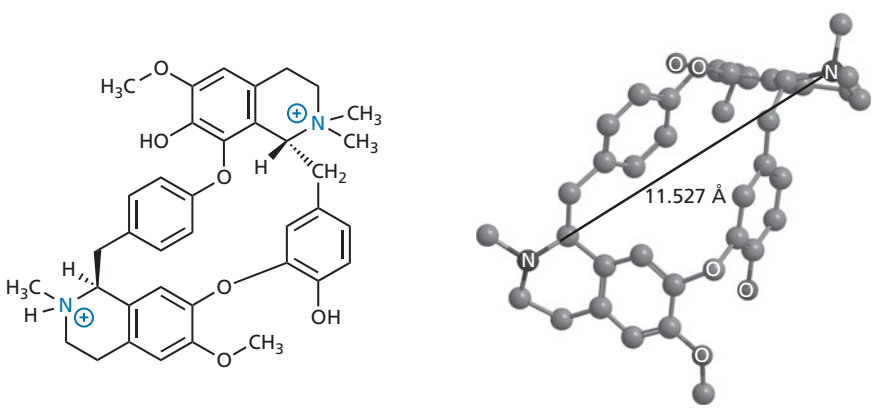


FIGURE 1 Computer-generated model of tubocurarine (Chem3D).

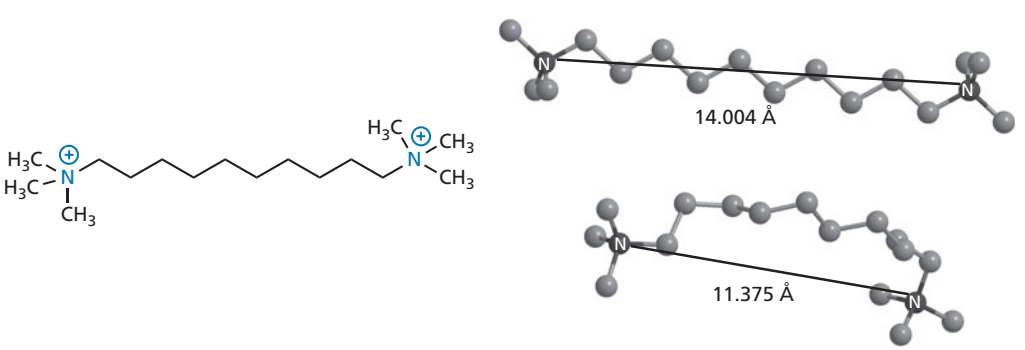


FIGURE 2 Computer-generated conformations of decamethonium (Chem3D).

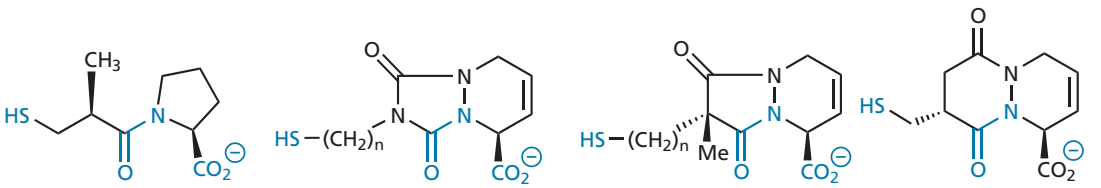


FIGURE 17.23 Captopril and rigid analogues (binding groups coloured).

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 antihypertensive agent **captopril** (Fig. 17.23). 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, rigid analogues (I–III) were synthesized, where the amide and the carboxyl group were fixed in space with respect to each other. Due to bond rotation, the thiol group can still access a sizeable volume. The biological activity of these compounds was then measured to identify which analogues were still active. A conformational analysis was then carried out as follows.

The possible conformations for captopril arising from bond rotation around the two bonds shown in Fig. 17.24 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. 17.25).

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

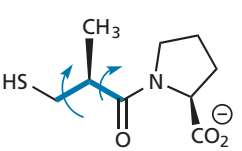


FIGURE 17.24 Bond rotations in captopril.

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

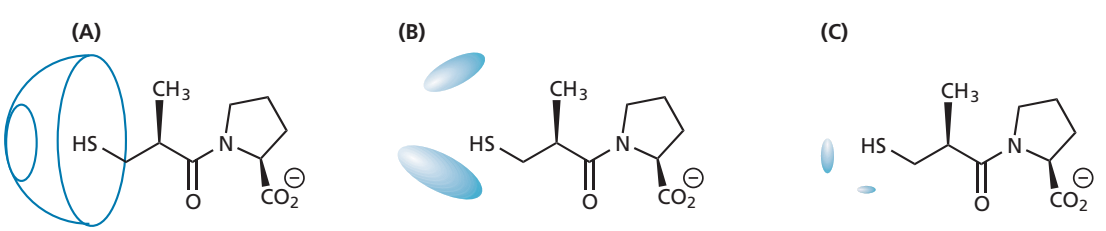


FIGURE 17.25 Generating spatial maps in conformational analysis. (**A**) All possible conformations, (**B**) stable conformations, and (**C**) after overlaps.

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 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. 17.26).

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 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

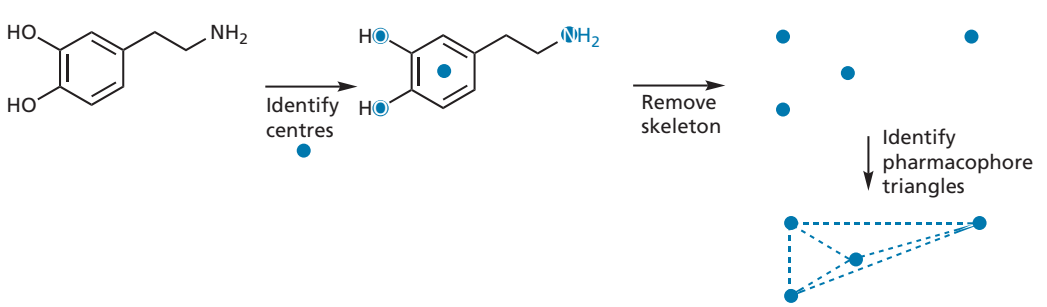


FIGURE 17.26 Pharmacophore identification in dopamine.

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 spotted quickly as they are clustered close together in specific regions of the plot.

For example, the three structures in Fig 17.27 were analysed and found to have 38 common pharmacophores. When these were plotted, seven distinct groups of pharmacophore were identified (Fig. 17.28). Each 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 docked successfully, 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.

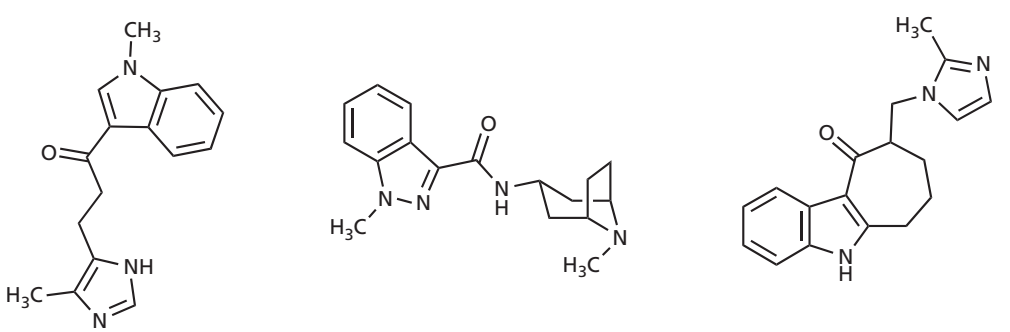


FIGURE 17.27 Test structures.

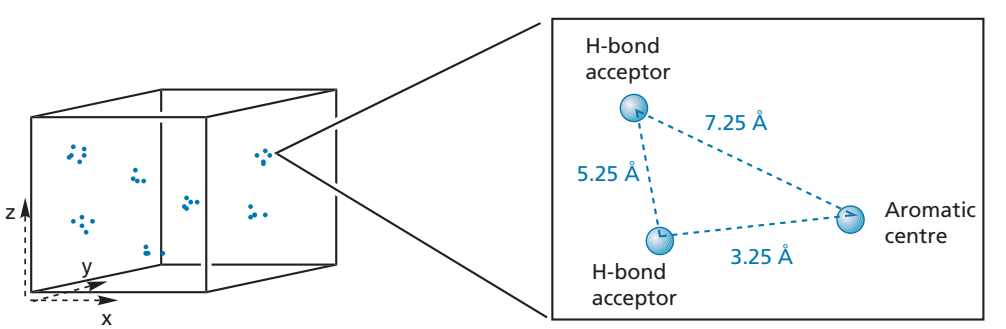


FIGURE 17.28 Pharmacophore plot.

17.12.2 Automatic docking

A variety of docking programs now exist 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 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 as 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 onto a 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. 17.29).

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 1.4–1.5 Å is used to represent a water molecule and this is 'rolled over' the surface of the binding site (Fig. 17.30). 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

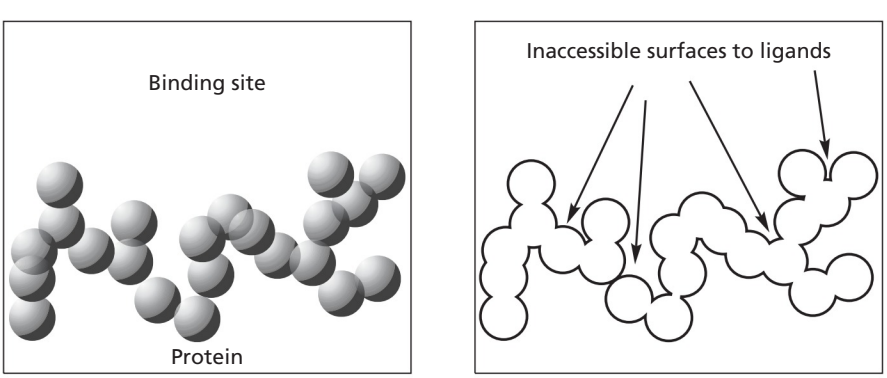


FIGURE 17.29 Defining the surface of the binding site by atoms and van der Waals surfaces.

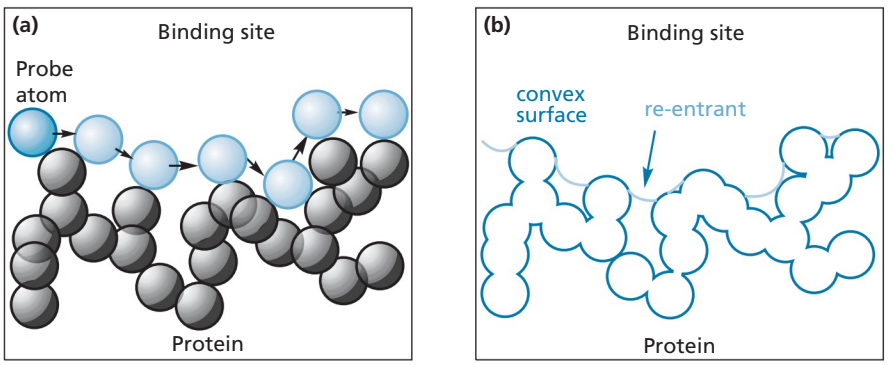


FIGURE 17.30 (**a**) Defining the Connolly surface of a binding site with a probe atom. (**b**) The Connolly surface shown in blue.

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 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. 17.31). This is achieved as follows.

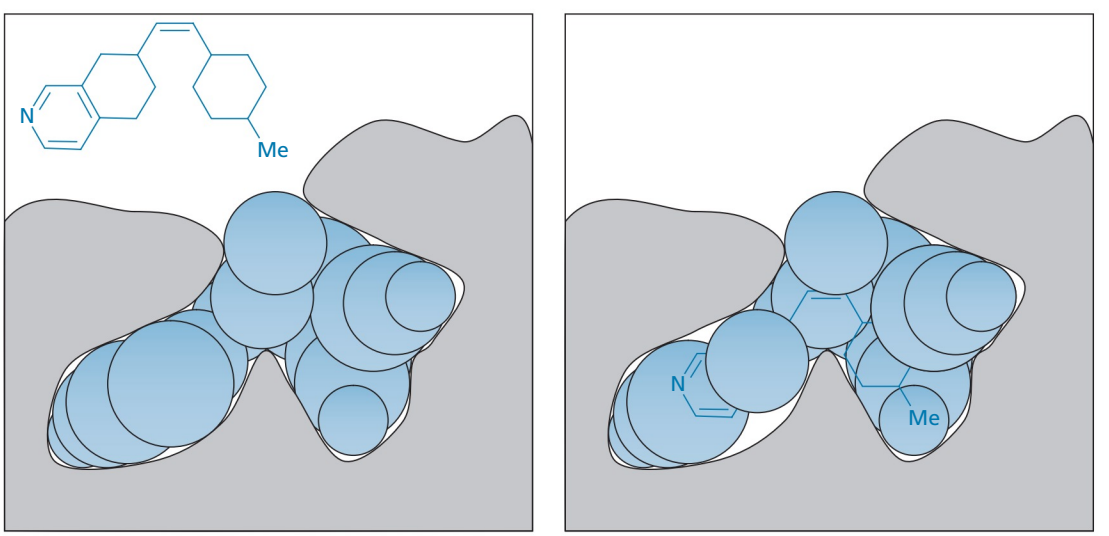


FIGURE 17.31 The DOCK program.

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 reduced significantly 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 pseudoatoms then overlaid. However, how does the programme decide which ligand atom and pseudoatom 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**. Firstly, the distances between each of the ligand atoms are measured. This is repeated for all of the pseudoatoms. These distances are then used to identify which ligand atoms and pseudoatoms can be matched up. The operation takes place as follows. 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/pseudoatoms (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 pseudoatoms

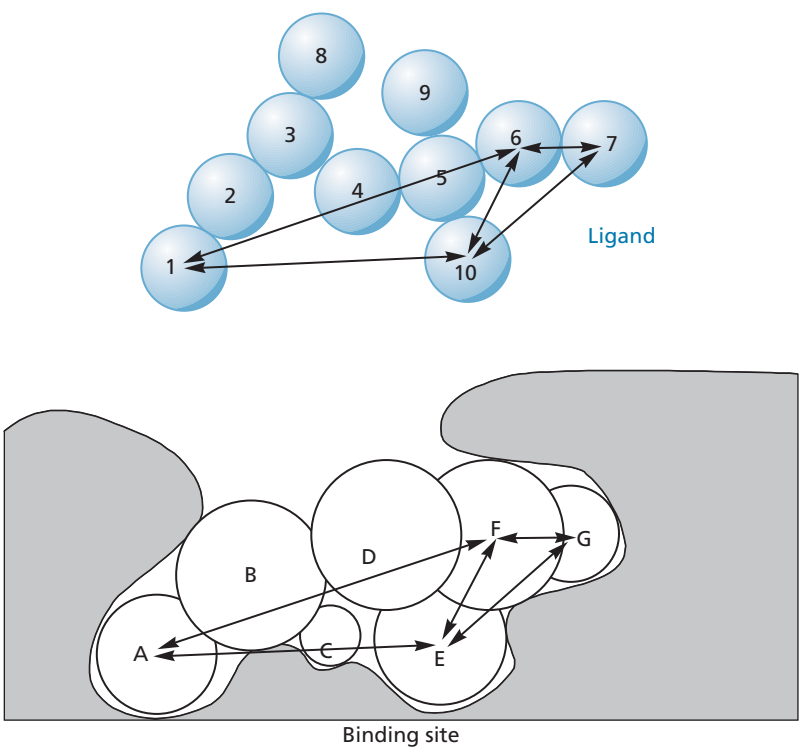


FIGURE 17.32 Comparing distances between ligand atoms and distances between pseudoatoms.

A–G (Fig. 17.32). The atom/pseudoatom pairs 1A, 6F, 7G, and 10E would be identified as compatible for the docking operation as the distances between the specified ligand atoms match the distances between the specified pseudoatoms.

Once this procedure has been carried out, the actual docking process can take place. Docking then involves an overlay where ligand atoms are fitted onto their paired pseudoatoms, as described in section 17.9. For example, in Fig. 17.33, ligand atoms 1, 6, 7, and 10 are matched to pseudoatoms A, F, G, and E respectively (Fig. 17.33). 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 atoms can match up with selected pseudoatoms). It takes no account of possible binding interactions, either favourable or unfavourable. Moreover, as selected ligand atoms are matched up with selected pseudoatoms, 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. 17.34). 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.

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.

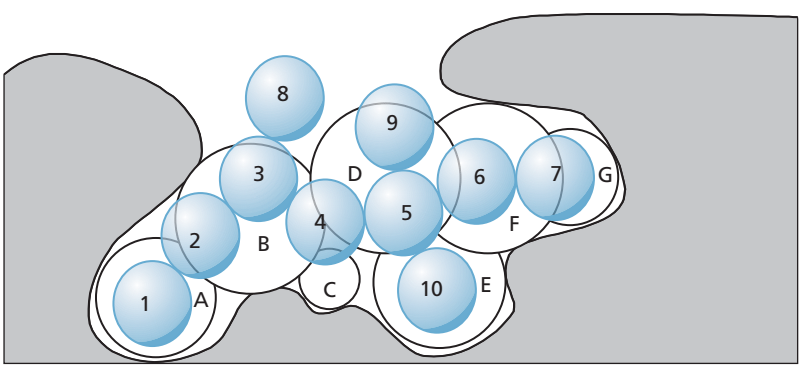


FIGURE 17.33 Docking by overlaying the atom–pseudoatom pairings of 1A, 6F, 7G, and 10E.

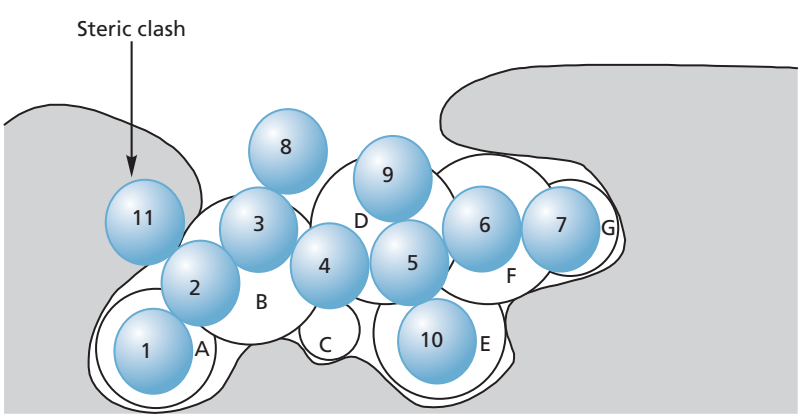


FIGURE 17.34 An unacceptable binding mode due to a steric clash.

17.12.5 The 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 accessed automatically for each atom of the ligand based on its particular position within the grid. This means that binding energies can be obtained speedily 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 with 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. Firstly, a hydrogen bonding group on the ligand must be the correct distance from a hydrogen bonding group in the binding site. Secondly, 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.

Firstly, a sphere is created around each hydrogen bonding group in the binding site (Fig. 17.35). 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 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 accessible 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. **Flexible Ligands Orientated on Grid** (**FLOG**) 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, which can lead to a **conformational explosion**. In other words, the number of

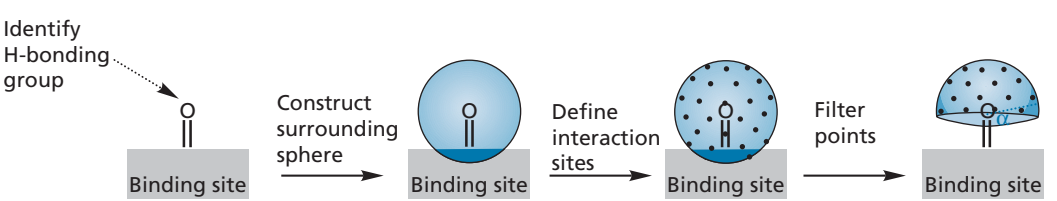


FIGURE 17.35 Identifying interaction sites for hydrogen bonding groups in the binding site.

possible conformations increases exponentially with the number of rotatable bonds present.

17.12.8 Docking of flexible ligands: anchor and grow programs

Various programs 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.

17.12.8.1 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. 17.36).

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. 17.36). At each stage of the process, energy minimization of the construct is carried out to relieve any strain arising from the construction process.

17.12.8.2 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 around each potential binding group in the binding site (Fig 17.37) (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 the ligand's pharmacophore triangles with the pharmacophore triangles present in the binding site. For a match to occur, a triangle for the ligand must have

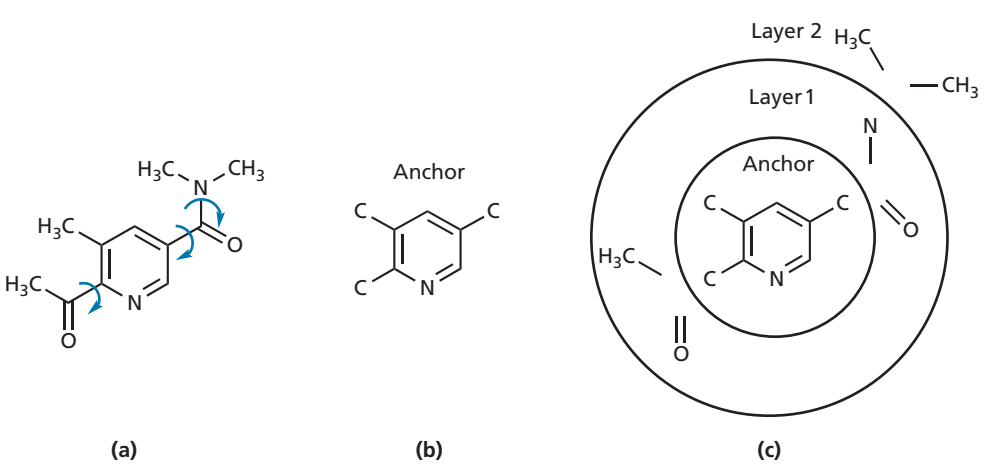


FIGURE 17.36 Anchor and grow algorithm. (**a**) Identify rotatable bonds. (**b**) Identify and dock a rigid anchor. (**c**) Add molecular fragments in layers.

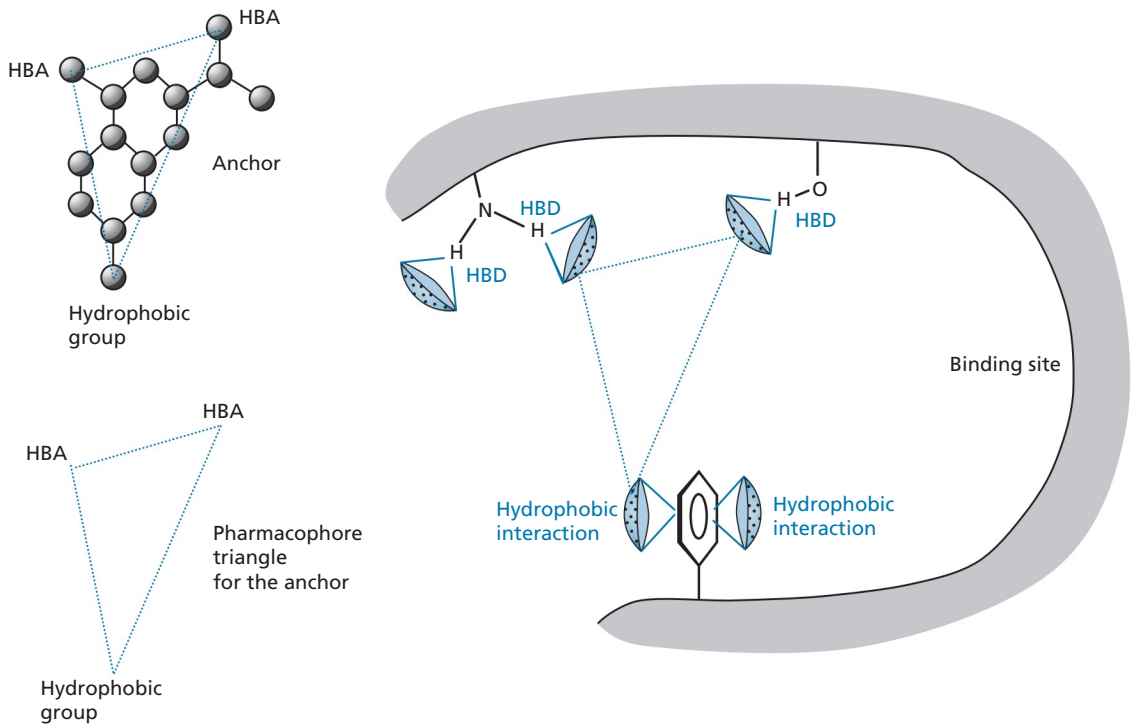


FIGURE 17.37 Docking based on the matching up of complementary pharmacophoric triangles for the anchor and the binding site (HBD = hydrogen bond donor; HBA = hydrogen bond acceptor).

roughly the same dimensions as a triangle in the binding site. Moreover, the corners of the triangles 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. 17.38).

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 as the program builds a set of interaction points round the anchor atoms prior to docking. For example, consider

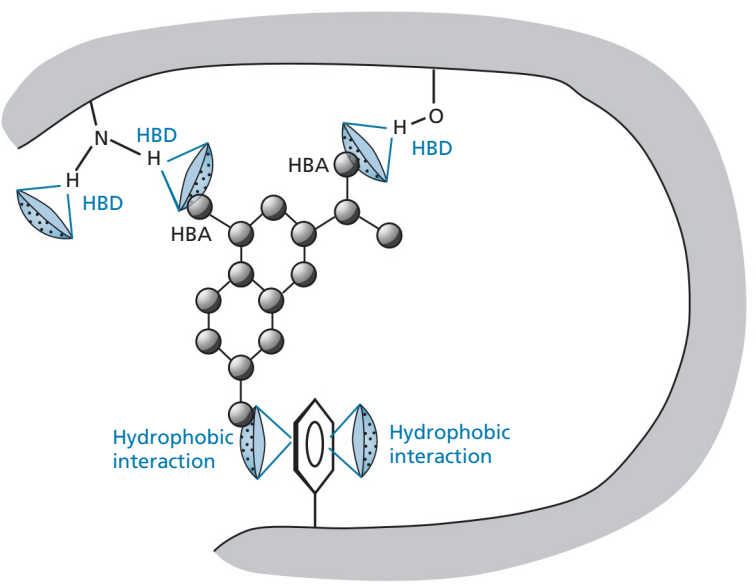


FIGURE 17.38 Docking of the anchor (HBD = hydrogen bond donor; HBA = hydrogen bond acceptor).

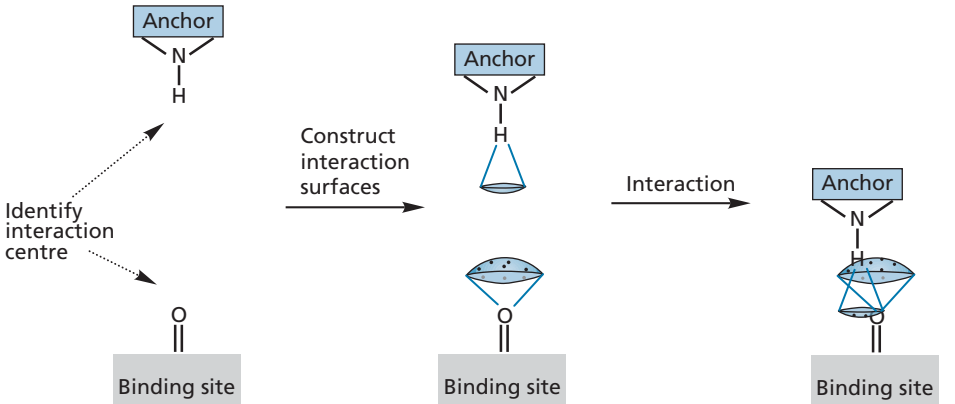


FIGURE 17.39 Assessing whether groups are at the correct distance and orientation with respect to each other for a good hydrogen bond interaction.

the case where docking matches up an N–H group on the anchor with a carbonyl group in the binding site (Fig. 17.39). 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 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 re-joining 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 that 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.

17.12.8.3 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 interactions (Fig. 17.40). The probes used are hydrogen atoms, as well as C=O and N–H fragments. Each of the probes can be scored as high-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

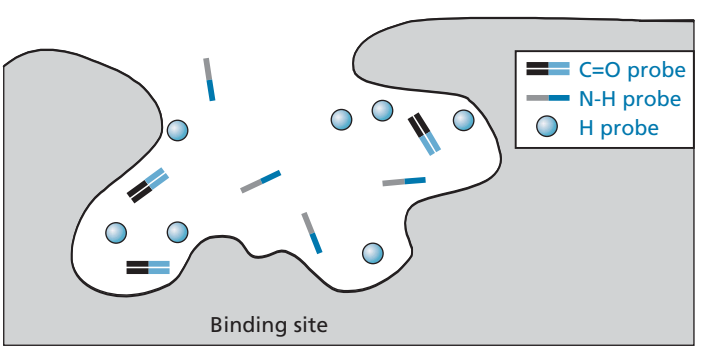


FIGURE 17.40 Positioning probes into a binding site to identify binding interactions.

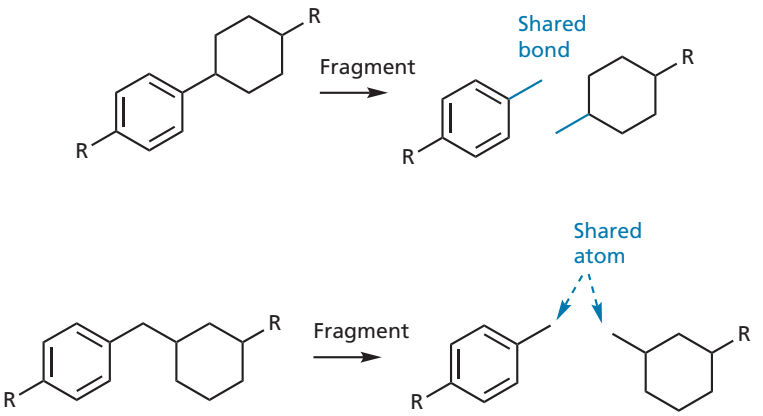


FIGURE 17.41 Fragmenting the ligand.

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. 17.41).

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 17.42 and 17.43). 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.

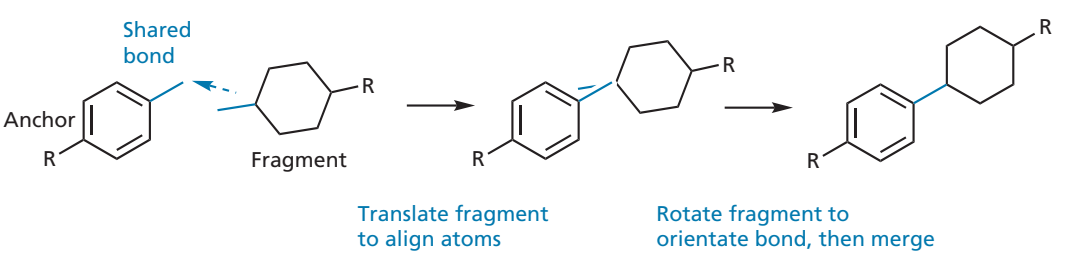


FIGURE 17.42 Merging fragments with a shared bond.

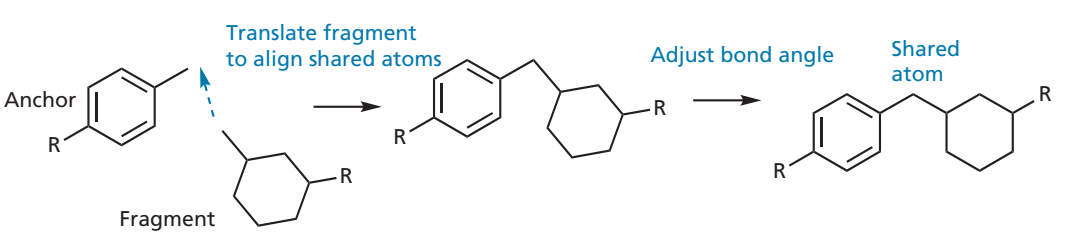


FIGURE 17.43 Merging fragments with a shared atom.

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 as the anchor with the highest binding score does not necessarily produce the best docking mode for the final ligand.

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 with 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. 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 prepared initially 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 change the molecule's conformation and orientation 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 binding 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-ray 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 these, model proteins and binding sites may be constructed to aid the drug design process.

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. 17.44). 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 revealed 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.

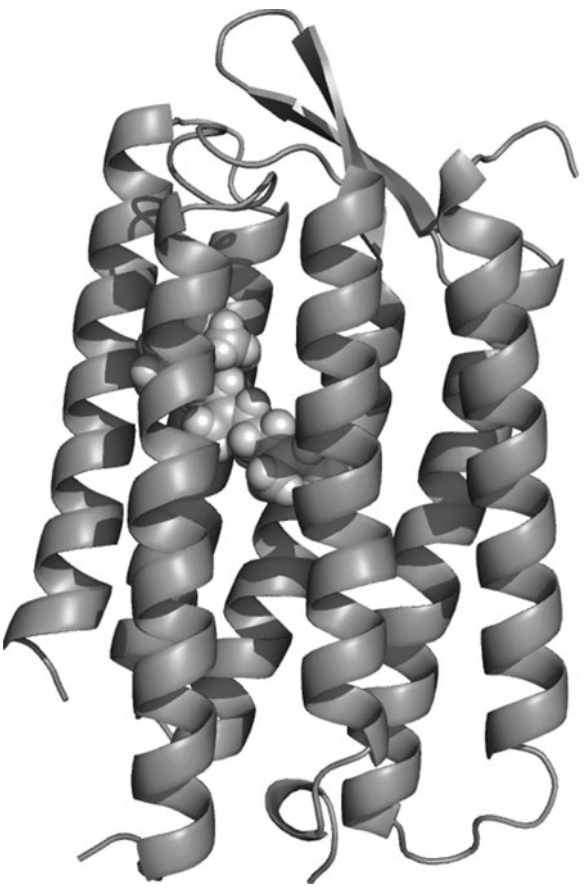


FIGURE 17.44 Bacteriorhodopsin with bound retinal.

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 both 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 these have identified amino acids which are important in binding neurotransmitters in a range of G-protein-coupled receptors (Fig. 17.45).

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

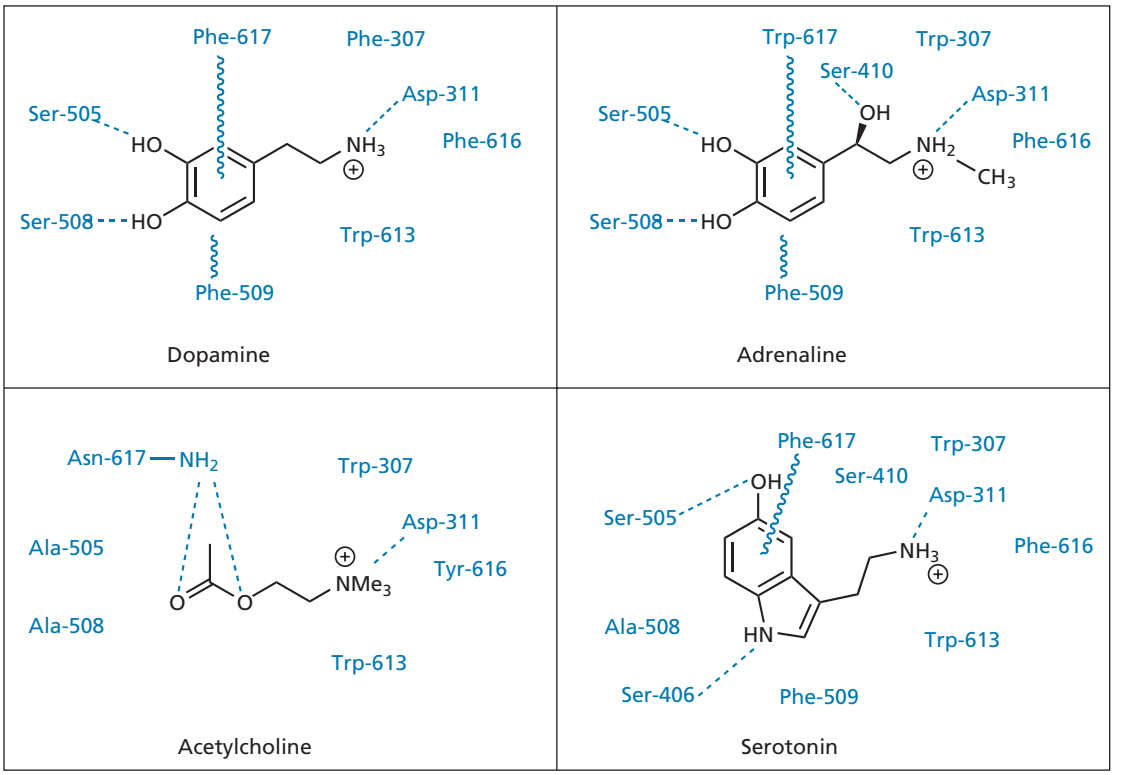


FIGURE 17.45 Important amino acids in various binding sites (the numbering indicates the position of each amino acid on the seven possible helices of a G-protein-coupled receptor, e.g. 311 indicates position 11 on helix 3).

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, which allows an interaction with the aromatic portion of the catecholamines, whereas in the cholinergic receptor this amino acid is asparagine, which allows 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 structurally different compounds with a range of activities is chosen. The active conformations are identified as far as possible and a 3D pharmacophore is 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 (see 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 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 phenylalanine, tryptophan, isoleucine, leucine, or valine could be used for a hydrophobic interaction (see also Box 17.5).

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

BOX 17.5 Constructing a receptor map

A range of structures, including **altanserin** and **ketanserin**, were used to construct a model receptor binding site for the 5-HT_{2a} receptor. Taking ketanserin as the representative structure for these compounds, various potential 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, along with 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.

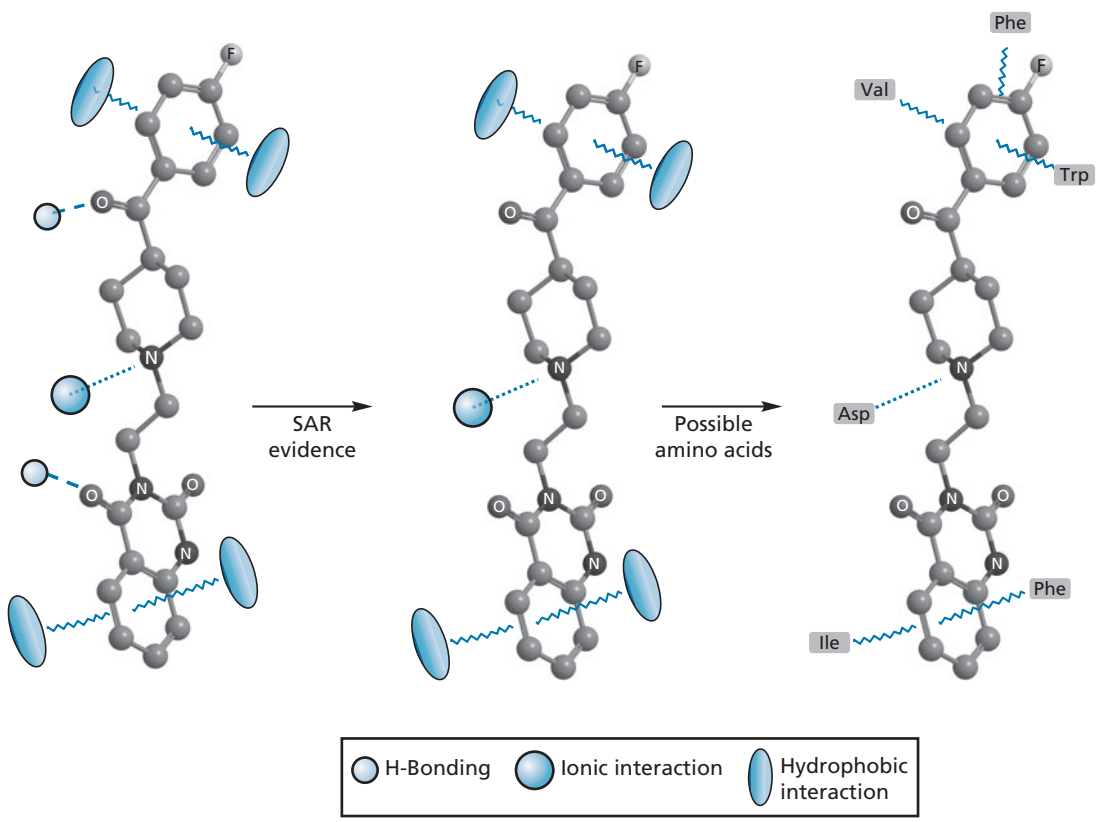


FIGURE 1 Receptor map for the 5-HT_{2a} receptor using ketanserin as the representative ligand.

section 18.10. Examples of specific software programs that generate hypothetical pseudoreceptors are 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 occupied by binding groups which 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* drug design

17.15.1 General principles of *de novo* drug design

De novo drug 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* drug 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 structure 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 several software packages 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 were identified regarding manual *de novo* drug design.

Firstly, 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 of 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 which are difficult or impossible to synthesize;
- similarly, it is pointless designing molecules which 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* drug design and some are particularly problematic. For example, automated *de novo* drug 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 drug has 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* drug design are 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 optimized through structure-based drug design (see, for example, section 20.7.4.4).

17.15.2 Automated *de novo* drug 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. 17.46). 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

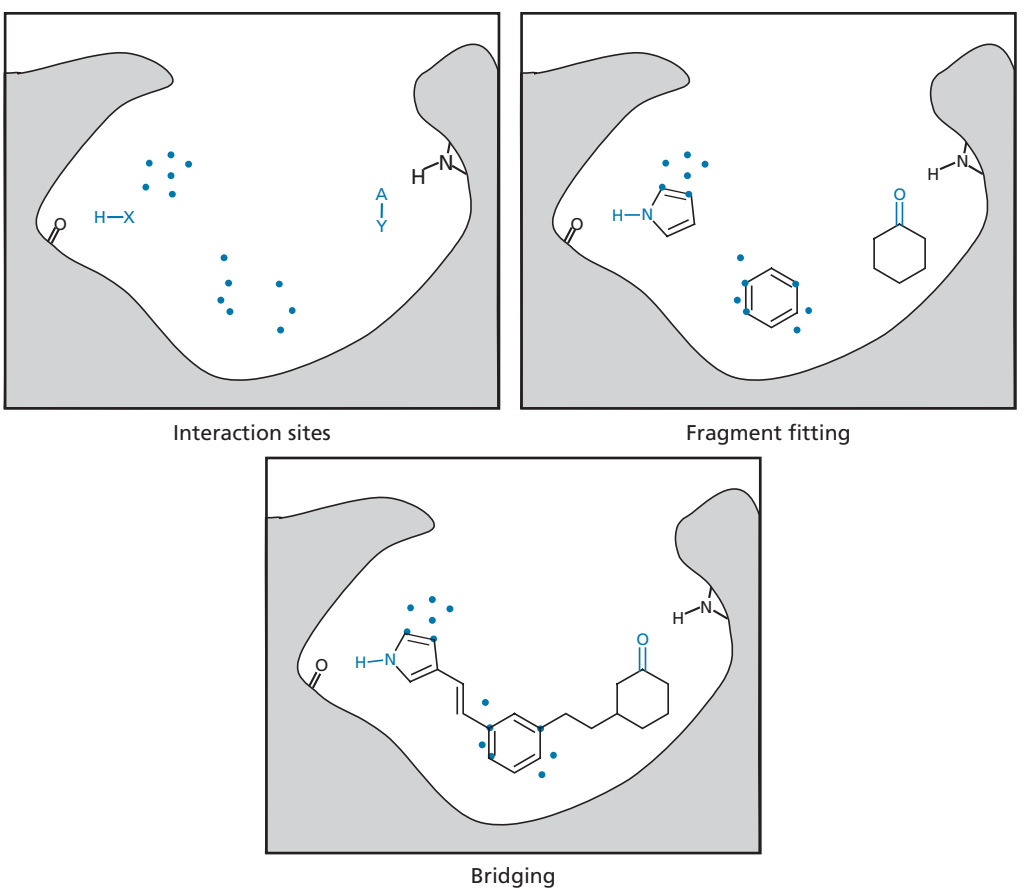


FIGURE 17.46 Stages involved in automated *de novo* drug design using LUDI (H–X = hydrogen bond donor interaction site; A–Y = hydrogen bond acceptor interaction site). The dots indicate aromatic interaction sites.

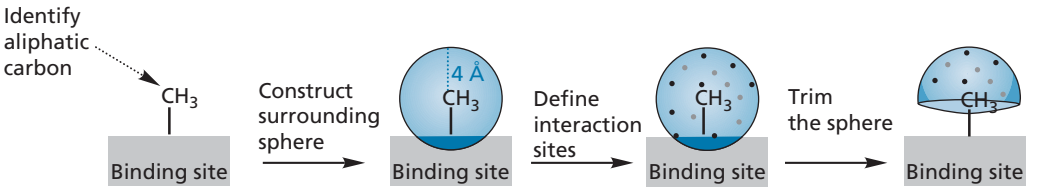


FIGURE 17.47 Identification of aliphatic interaction sites around a methyl group (LUDI).

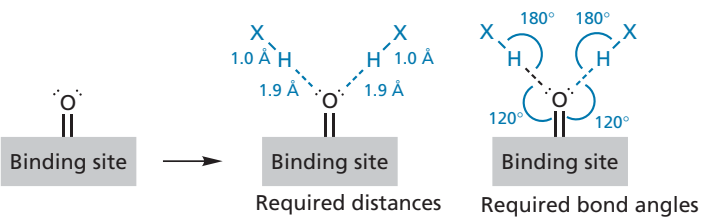


FIGURE 17.48 The interaction sites for a hydrogen bond donor, represented by H–X (LUDI).

atoms are identified as hydrogen bond acceptors. Any 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 that define 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. 17.47). 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 also the relevant orientation of the atoms. 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) which can be determined (Fig. 17.48).

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. 17.49 shows how the interaction site for a hydrogen 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. 17.50).

As with the van der Waals interaction sites, hydrogen bond interaction sites are checked to ensure that they are

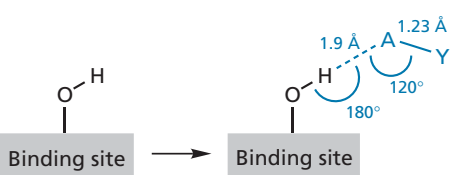


FIGURE 17.49 The interaction site for a hydrogen bond acceptor represented by A–Y (LUDI). A denotes the hydrogen bond acceptor.

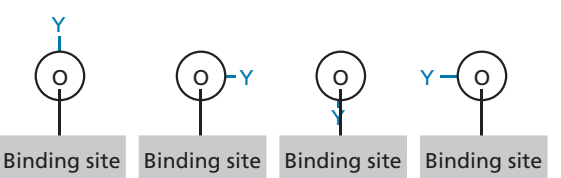


FIGURE 17.50 Four possible interaction sites for A–Y (A is hidden).

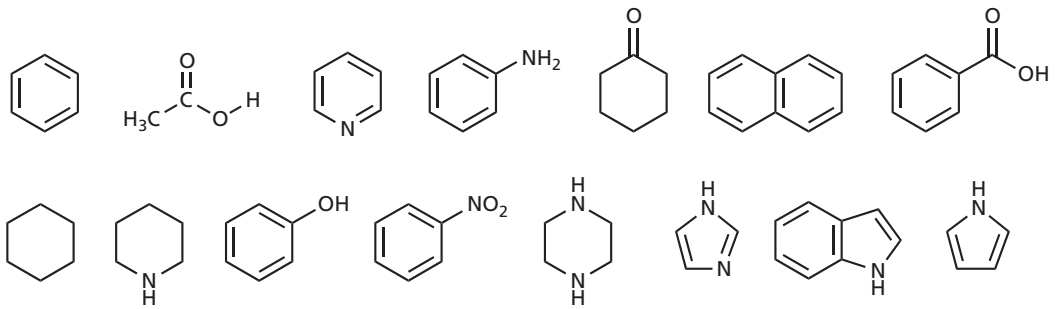


FIGURE 17.51 Examples of molecular fragments used by LUDI.

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. 17.51. 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 onto the corresponding interaction site (Fig. 17.52). 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

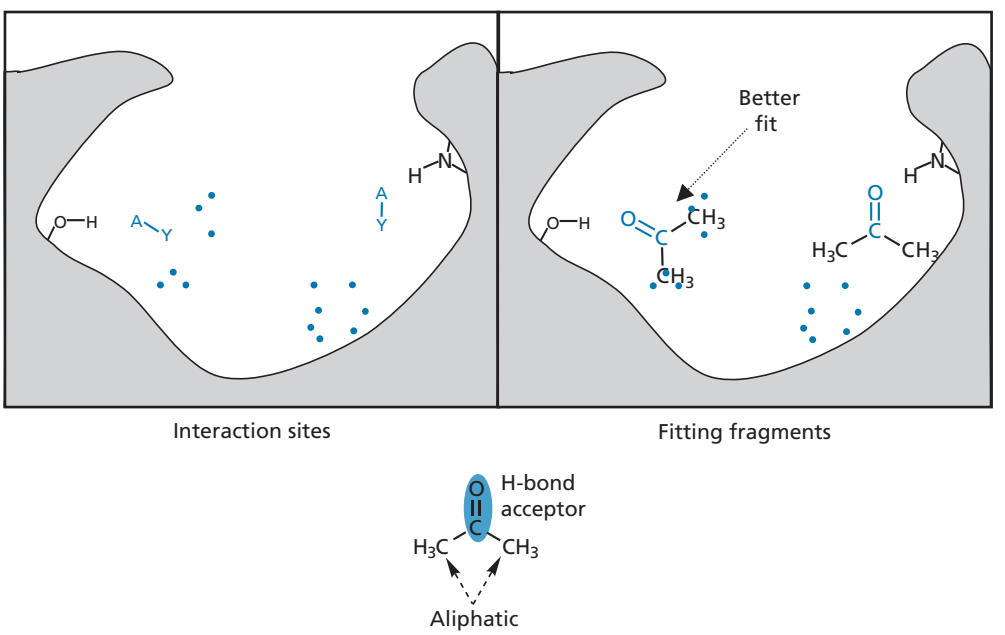


FIGURE 17.52 Fitting fragments (A–Y = hydrogen bond acceptor interaction site). The dots indicate aliphatic interaction sites.

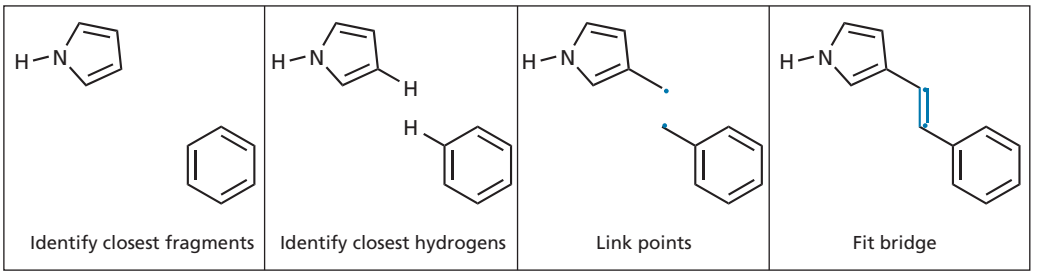


FIGURE 17.53 The bridging process (LUDI).

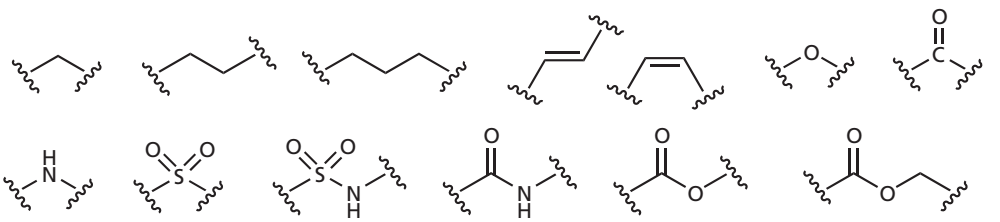


FIGURE 17.54 Examples of molecular bridges (LUDI).

closest to each other in the binding site, then identifies the closest hydrogen atoms (Fig 17.53). 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. 17.54. 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 program 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 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 **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. Templates, however, are designed to represent several different molecular fragments. Each template is defined by vertices and edges, rather than by atoms and bonds. A vertex represents a generalized 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 template shown in Fig. 17.55 can represent a large number of different six-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 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. 17.56). In the first stage, the emphasis is on

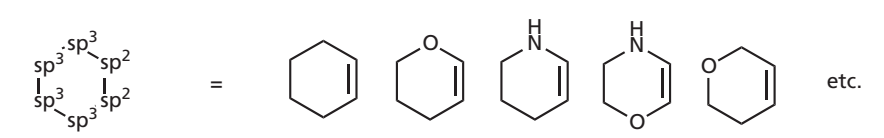


FIGURE 17.55 Examples of structures represented by a template used in SPROUT.

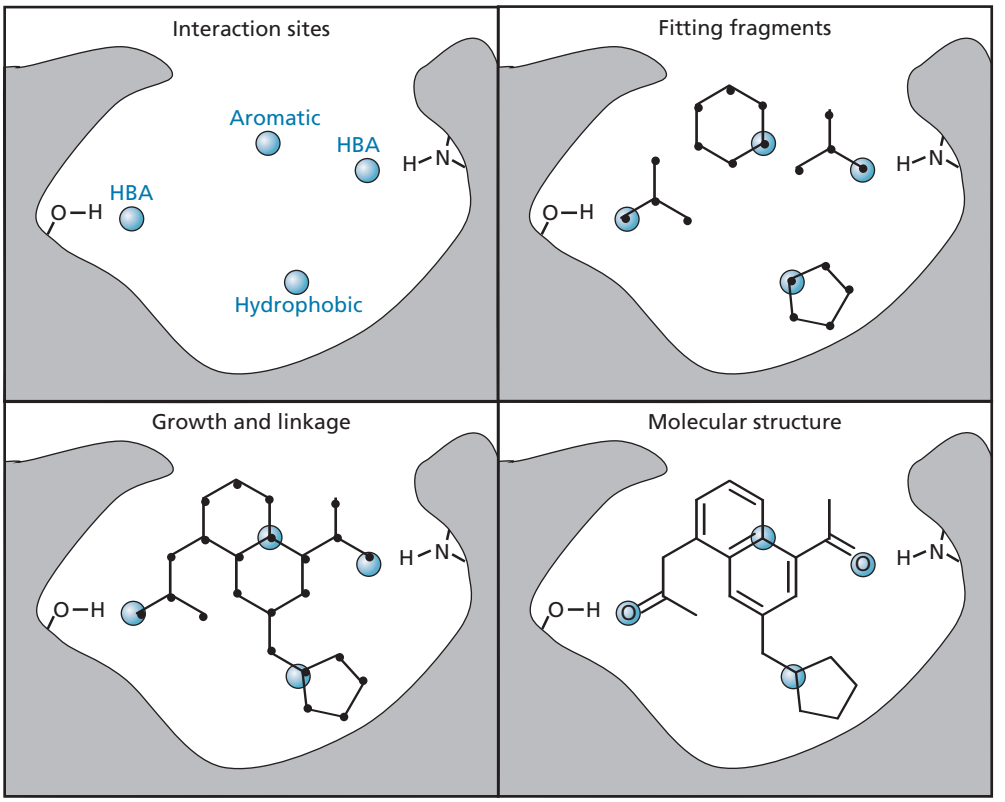


FIGURE 17.56 Generating structures using SPROUT.

generating fragment 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 fragment templates were then added sequentially, and the skeleton was 'grown' until it occupied all the other spheres. In the current version, fragment templates 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 fragments 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 specific molecules from the molecular templates 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. As 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 tautomerize 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. 17.57).

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. 17.58). In this example, the link could be made synthetically by reacting an alkoxide with an alkyl halide (Fig. 17.59).

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

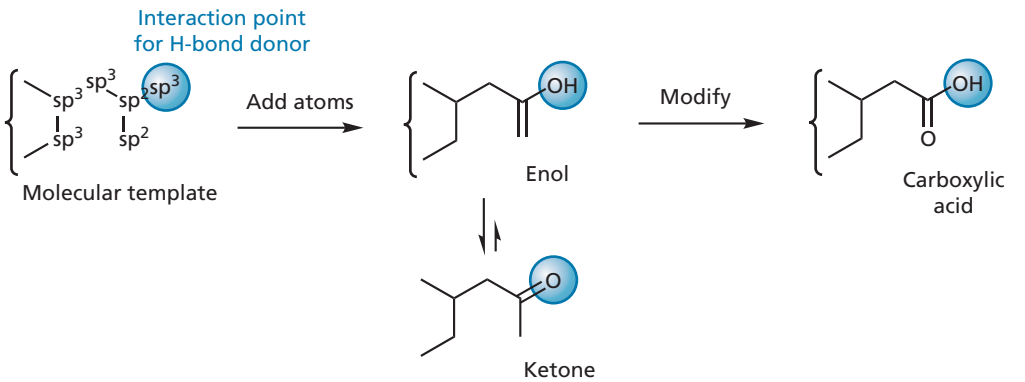


FIGURE 17.57 Modification of an enol to a carboxylic acid by SPROUT.

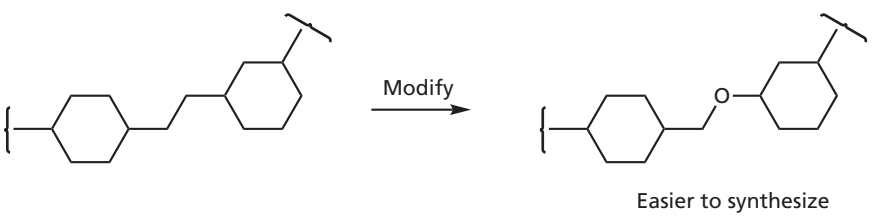


FIGURE 17.58 Modification by SPROUT to generate a more synthetically feasible structure.

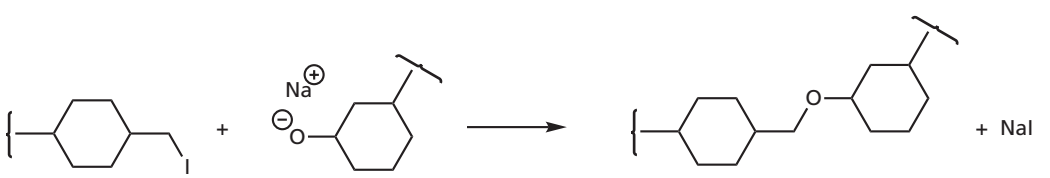


FIGURE 17.59 Possible synthesis allowing the linkage required in Fig. 17.58.

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 as 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 as 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 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 variously 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 10 most frequent substitution patterns for a naphthalene ring amongst a database of known compounds are shown in Fig. 17.60. The analysis of partial structures can be carried out such that structures with uncommon structural features (such as a tetrasubstituted naphthalene ring) are penalized and rejected. A measure of the drug-like character of the partial structures can also be gleaned if the database used in the analysis is restricted to active compounds from drug databases.

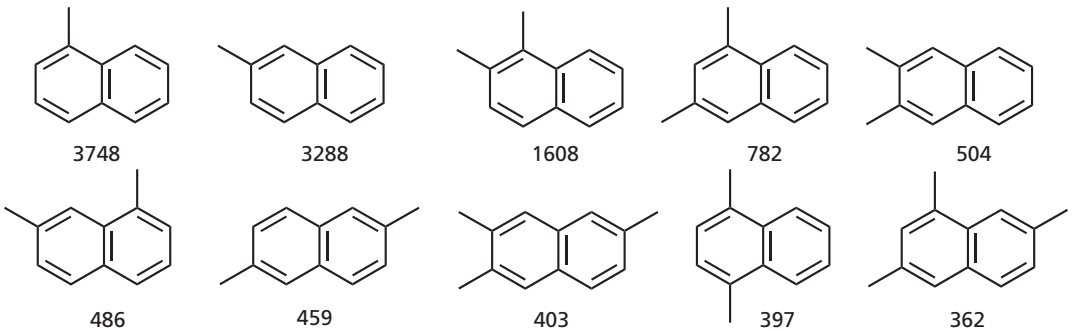


FIGURE 17.60 The 10 most frequent substitution patterns for naphthalene in a database of drug-like compounds. Numbers refer to number of occurrences.

17.15.2.3 LEGEND

LEGEND is another long-established automated *de novo* drug 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 as 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. Inevitably, this means that the generation of structures is a slower process compared with 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 predetermined 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 optimized, 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.

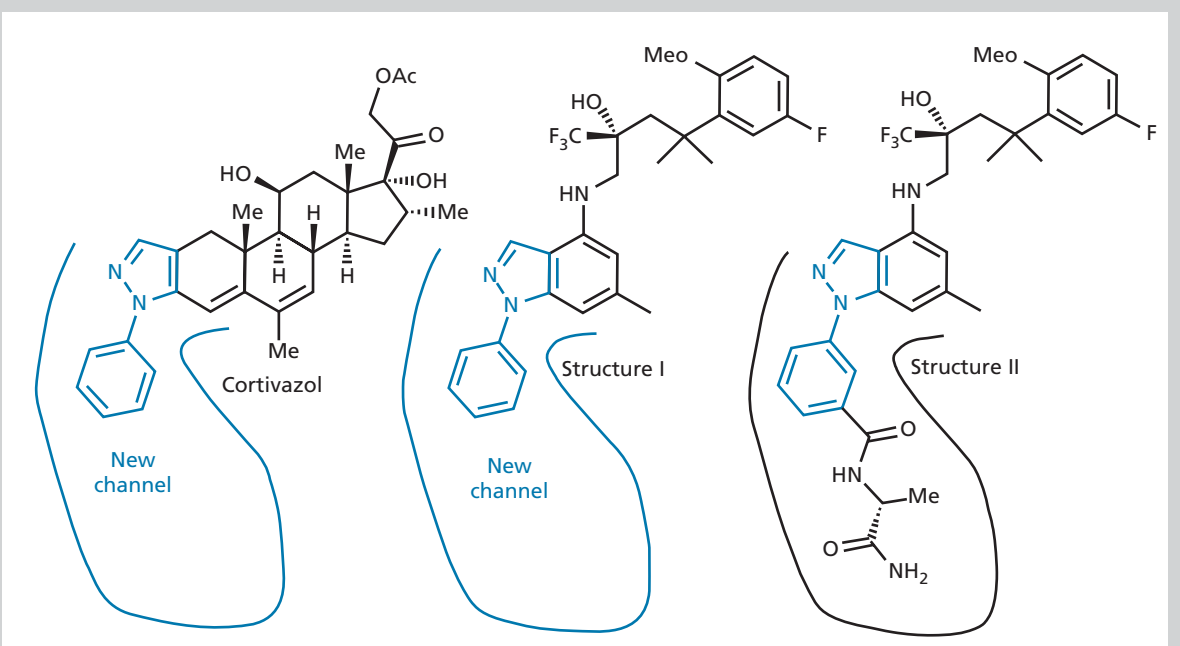
SYNOPSIS is a *de novo* drug design program that is designed to generate synthetically feasible structures. It does so by incorporating synthetic rules into the structure building process. In other words, fragments can only be linked if there is a known reaction which will allow it. Moreover, the fragments 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 with 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 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 d*e novo* design identify interaction sites in the binding site, then match molecular fragments to these sites. Bridges are then designed to link the fragments.

BOX 17.6 Designing a non-steroidal glucocorticoid agonist

The arylindazole structure (I) is a non-steroidal agent that acts as an agonist at the glucocorticoid receptor, but it is not possible to dock the structure into the conventional receptor binding site. However, an unusual induced fit has been observed for the steroid **cortivazol**, which opens up a new channel in the binding site (Box 8.1). Docking experiments of structure I with this atypical binding site were successful and it was found that the arylpyrazole group (shown in blue) partially occupied the new channel. Using structure I as a core scaffold a computational technique called **AlleGrow** was used to 'grow' the structure into the new channel *in silico*. The program works by adding atoms or small molecular features to the core skeleton, then scoring the resulting structures for binding interactions. Seven thousand virtual structures were created *in silico* by this method, and the most promising of these were synthesized and tested, leading to structure II. This was found to have similar activity to the most potent of the clinically used corticosteroids.



17.16 Planning compound libraries

Combinatorial and parallel synthesis (Chapter 16) are methods of rapidly creating a large number of compounds on a small scale using a set reaction scheme. The compounds produced constitute a **compound library** which could be tested to find active compounds for a set target. A compound library could be created that would include all the possible compounds obtainable from the reaction scheme using available starting materials and reagents. However, molecular modelling can help to focus the study such that a smaller number of structures are made, while maintaining the probability of finding active compounds.

One method of doing this is based on the identification of **pharmacophore triangles** (section 17.11.3). Let us assume that a 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 a similar size which produces only 100 different pharmacophores. Therefore, an effective way of designing a more focused library is to carry out a pharmacophore search on all the possible products from a reaction scheme, in order to select those products that demonstrate the widest structural diversity. Those compounds would then be the ones chosen for inclusion in the library.

Firstly, all the possible synthetic products are ranked automatically 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 and all the possible pharmacophore triangles are identified for that 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 compound 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 preselected, 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 large amounts of data. For example, activity against a range

of targets has to be measured to ensure that the compounds have good activity against their intended target, and also show 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, a 2D plot could be drawn up comparing α-adrenergic activity and 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.

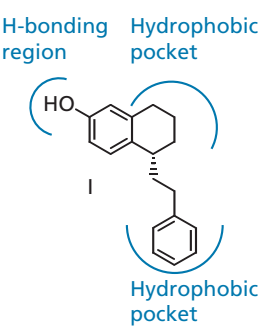
Some programs can be used to assess the biological results from a compound library. Firstly, the scaffold used in the library is defined, then the substituents are defined. Once the biological test results are obtained, a tree diagram can be drawn up to assess which substitution point is most important 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.

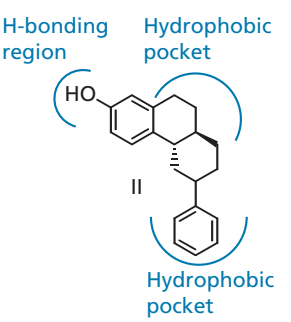
KEY POINTS

- 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 with 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 is their relevance 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 *de novo* drug design to find a ligand for a binding site that contains a hydrogen bonding region and two hydrophobic pockets. Structures I and II are both suitable candidates. Compare the relevant merits of these structures and decide which one you would synthesize first to test your binding theory.

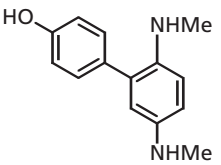




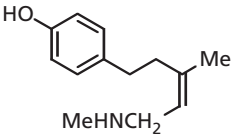
6. Both structures I and II show poor water solubility. It is suggested that the phenyl group be replaced by a pyridine ring. What would be the advantages and disadvantages of this idea? Have you any alternative ideas?
7. Assuming that structures I and II both bind to the binding site as predicted, what further modifications might you make to increase binding interactions?

8. Why were such modifications not carried out earlier?

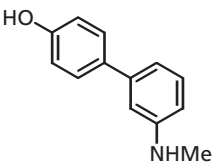
9. The following eight structures have been tested for activity as receptor agonists. Five are active and three are inactive. Assess the structures and discuss what the pharmacophore might be for agonist activity.



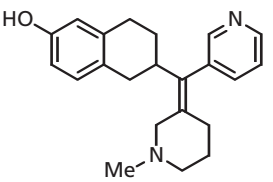
Active



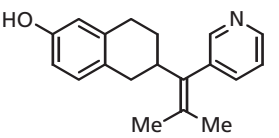
Active



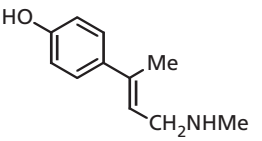
Active



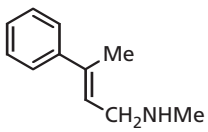
Active



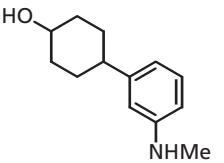
Inactive



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.

Biggadike, K., Bledsoe, R. K., Coe, D. M., Cooper, T. W., House, D., Iannone, M. A., et al. (2009) Design and x-ray crystal structures of high potency nonsteroidal glucocorticoid agonists exploiting a novel binding site on the receptor. *Proceedings of the New York Academy of Sciences* **106**, 18114–18119.

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 Review of Biophysics and Biomolecular Structure* **32**, 335–373.

Ganellin, C. R. and Roberts, S. M. (eds) (1994) Discovery and development of cromokalim and related potassium channel activators. 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., Decornez, H., Furr, J. R., and Bajorath, J. 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.

Megget, K. (2011) Idle cures. *Chemistry World* February, 52–55.

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.

Sansom, C. (2010) Model molecules. *Chemistry World* April, 50–53.

Sansom, C. (2010) Receptive receptors. *Chemistry World* August, 52–55.

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. 763.

18 Quantitative structure–activity relationships (QSAR)

In Chapters 13 and 14 we studied the various strategies that can be used in the design of drugs. Several of these strategies involved a change in shape such that the new drug had a better 'fit' for its target binding site. Other strategies involved a change in functional groups or substituents such that the drug's pharmacokinetics or binding site interactions were improved. These latter strategies often involved the synthesis of analogues containing a range of substituents on aromatic or heteroaromatic rings or accessible functional groups. The number of possible analogues that could be made is infinite if we were to try and synthesize analogues with every substituent and combination of substituents possible. Therefore, it is clearly advantageous if a rational approach can be followed in deciding which substituents to use. The quantitative structure–activity relationship (QSAR) approach has proved extremely useful in tackling this problem.

The QSAR approach attempts to identify and quantify the physicochemical properties of a drug and to see whether any of these properties has an effect on the drug's biological activity. If such a relationship holds true, an equation can be drawn up which quantifies the relationship and allows the medicinal chemist to say with some confidence that the property (or properties) has an important role in the pharmacokinetics or mechanism of action of the drug. It also allows the medicinal chemist some level of prediction. By quantifying physicochemical properties, it should be possible to calculate in advance what the biological activity of a novel analogue might be. There are two advantages to this. Firstly, it allows the medicinal chemist to target efforts on analogues which should have improved activity and, thus, cut down the number of analogues that have to be made. Secondly, if an analogue is discovered which does not fit the equation, it implies that some other feature is important and provides a lead for further development.

What are these physicochemical features that we have mentioned?

Essentially, they could be any structural, physical, or chemical property of a drug. Clearly, any drug will have a large number of such properties and it would be a Herculean task to quantify and relate them all to biological activity at the same time. A simple, more practical approach is to consider one or two physicochemical properties of the drug and to vary these while attempting to keep other properties constant. This is not as simple as it sounds, as it is not always possible to vary one property without affecting another. Nevertheless, there have been numerous examples where the approach has worked.

It is important that the QSAR method is used properly and in relevant situations. Firstly, the compounds studied must be structurally related, act at the same target, and have the same mechanism of action. Secondly, it is crucial that the correct testing procedures are used. *In vitro* tests carried out on isolated enzymes are relevant for a QSAR study as the activities measured for different inhibitors are related directly to how each compound binds to the active site. *In vivo* tests carried out to measure the physiological effects of enzyme inhibitors are not valid, however, as both pharmacodynamic and pharmacokinetic factors come into play. This makes it impossible to derive a sensible QSAR equation.

18.1 Graphs and equations

In the simplest situation, a range of compounds is synthesized in order to vary one physicochemical property (e.g. log P) and to test how this affects the biological activity (log $1/C$) (we will come to the meaning of log $1/C$ and log P in due course). A graph is then drawn to plot the biological activity on the y-axis versus the physicochemical feature on the x-axis (Fig. 18.1).

It is then necessary to draw the best possible line through the data points on the graph. This is done by a procedure known as **linear regression analysis by the least squares method**. This is quite a mouthful and can produce a glazed expression on any chemist who is not mathematically orientated. In fact, the principle is quite straightforward. If we draw a line through a set of data

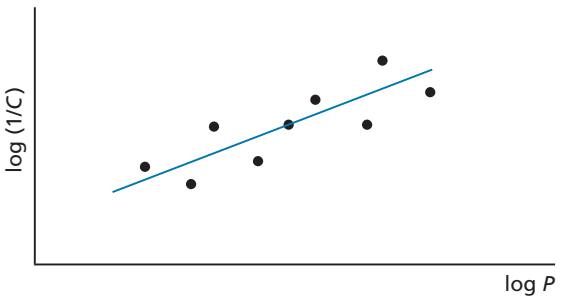


FIGURE 18.1 Biological activity versus log *P*.

points, most of the points will be scattered on either side of the line. The best line will be the one closest to the data points. To measure how close the data points are, vertical lines are drawn from each point (Fig. 18.2). These verticals are measured and then squared in order to eliminate the negative values. The squares are then added up to give a total (the sum of the squares). The best line through the points will be the line where this total is a minimum. The equation of the straight line will be $y = k_1x + k_2$, where k_1 and k_2 are constants. By varying k_1 and k_2, different equations are obtained until the best line is obtained. This whole process can be done speedily using relevant software.

The next stage in the process is to see whether the relationship is meaningful. As any good politician knows, you can make figures mean anything and so statistical evidence has to be obtained to support the QSAR equation and to quantify the goodness of fit. Otherwise, we may have obtained a straight line through points which are so random that it means nothing. The **regression** or **correlation coefficient** (r) is a measure of how well the physicochemical parameters present in the equation explain the observed variance in activity. An explanation of how r is derived is given in Appendix 3. For a perfect fit $r = 1$, in which case the observed activities would be the same as those calculated by the equation. Such perfection is impossible with biological data and so r values greater than 0.9 are considered acceptable.

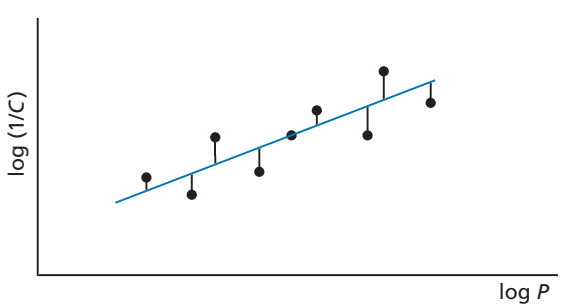


FIGURE 18.2 Proximity of data points to line of best fit.

The regression coefficient is often quoted as r^2, in which case values over 0.8 are considered a good fit. If r^2 is multiplied by 100 it indicates the percentage variation in biological activity that is accounted for by the physicochemical parameters used in the equation. Thus, an r^2 value of 0.85 signifies that 85% of the variation in biological activity is accounted for by the parameters used. There are dangers in putting too much reliance on r, as the value obtained takes no account of the number of compounds (n) involved in the study and it is possible to obtain higher values of r by increasing the number of compounds tested.

Therefore, another statistical measure for the goodness of fit should be quoted alongside r. This is the **standard error of estimate** or the **standard deviation** (s). Ideally, s should be zero, but this would assume there were no experimental errors in the experimental data or the physicochemical parameters. In reality, s should be small, but not smaller than the standard deviation of the experimental data. It is therefore necessary to know the latter to assess whether the value of s is acceptably low. Appendix 3 shows how s is obtained and demonstrates that the number of compounds (n) in the study influences the value of s.

Statistical tests called **Fisher's *F*-tests** are often quoted (Appendix 3). These tests are used to assess the significance of the coefficients k for each parameter in the QSAR equation. Normally, p values (derived from the F-test) should be less than or equal to 0.05 if the parameter is significant. If this is not the case, the parameter should not be included in the QSAR equation.

18.2 Physicochemical properties

Many physical, structural, and chemical properties have been studied by the QSAR approach, but the most common are hydrophobic, electronic, and steric properties. This is because it is possible to quantify these effects. Hydrophobic properties can be easily quantified for complete molecules or for individual substituents. However, it is more difficult to quantify electronic and steric properties for complete molecules, and this is only really feasible for individual substituents.

Consequently, QSAR studies on a variety of totally different structures are relatively rare and are limited to studies on hydrophobicity. It is more common to find QSAR studies being carried out on compounds of the same general structure, where substituents on aromatic rings or accessible functional groups are varied. The QSAR study then considers how the hydrophobic, electronic, and steric properties of the substituents affect biological activity. The three most studied physicochemical properties are now considered in some detail.

18.2.1 Hydrophobicity

Test your understanding and practise your molecular modelling with Exercise 18.1. you might also find exercises 25.3 and 25.4 useful at this point.

The hydrophobic character of a drug is crucial to how easily it crosses cell membranes (section 11.3) and may also be important in receptor interactions. Changing substituents on a drug may well have significant effects on its hydrophobic character and, hence, its biological activity. Therefore, it is important to have a means of predicting this quantitatively.

18.2.1.1 The partition coefficient (*P*)

The hydrophobic character of a drug can be measured experimentally by testing the drug's relative distribution in an *n*-octanol/water mixture. Hydrophobic molecules will prefer to dissolve in the *n*-octanol layer of this two-phase system, whereas hydrophilic molecules will prefer the aqueous layer. The relative distribution is known as the partition coefficient (*P*) and is obtained from the following equation:

$$P = \frac{\text{Concentration of drug in octanol}}{\text{Concentration of drug in aqueous solution}}$$

Hydrophobic compounds have a high *P* value, whereas hydrophilic compounds have a low *P* value.

Varying substituents on the lead compound will produce a series of analogues having different hydrophobicities and, therefore, different *P* values. By plotting these *P* values against the biological activity of these drugs, it is possible to see if there is any relationship between the two properties. The biological activity is normally expressed as 1/*C*, where *C* is the concentration of drug required to achieve a defined level of biological activity. The reciprocal of the concentration (1/*C*) is used, as more active drugs will achieve a defined biological activity at lower concentration.

The graph is drawn by plotting log (1/*C*) versus log *P*. In studies where the range of the log *P* values is restricted to a small range (e.g. log *P* = 1–4), a straight-line graph is obtained (Fig. 18.1) showing that there is a relationship between hydrophobicity and biological activity. Such a line would have the following equation:

$$\log\left(\frac{1}{C}\right) = -k_1 \log P + k_2$$

For example, the binding of drugs to serum albumin is determined by their hydrophobicity and a study of 42 compounds resulted in the following equation:

$$\log\left(\frac{1}{C}\right) = 0.75 \log P + 2.30 \quad (n = 42, r = 0.960, s = 0.159)$$

The equation shows that serum albumin binding increases as log *P* increases. In other words, hydrophobic drugs bind more strongly to serum albumin than hydrophilic drugs. Knowing how strongly a drug binds to serum albumin can be important in estimating effective dose levels for that drug. When bound to serum albumin, the drug cannot bind to its receptor and so the dose levels for the drug should be based on the amount of unbound drug present in the circulation. The equation above allows us to calculate how strongly drugs of similar structure will bind to serum albumin and gives an indication of how 'available' they will be for receptor interactions. The *r* value of 0.96 is close to 1, which shows that the line resulting from the equation is a good fit. The value of r^2 is 92%, which indicates that 92% of the variation in serum albumin binding can be accounted for by the different hydrophobicities of the drugs tested. This means that 8% of the variation is unaccounted for, partly as a result of the experimental errors involved in the measurements.

Despite such factors as serum albumin binding, it is generally found that increasing the hydrophobicity of a lead compound results in an increase in biological activity. This reflects the fact that drugs have to cross hydrophobic barriers, such as cell membranes, in order to reach their target. Even if no barriers are to be crossed (e.g. in *in vitro* studies), the drug has to interact with a target system, such as an enzyme or receptor, where the binding site is more hydrophobic than the surface. Therefore, increasing hydrophobicity also aids the drug in binding to its target site.

This might imply that increasing log *P* should increase the biological activity *ad infinitum*. In fact, this does not happen. There are several reasons for this. For example, the drug may become so hydrophobic that it is poorly soluble in the aqueous phase. Alternatively, it may be 'trapped' in fat depots and never reach the intended site. Finally, hydrophobic drugs are often more susceptible to metabolism and subsequent elimination.

A straight-line relationship between log *P* and biological activity is observed in many QSAR studies because the range of log *P* values studied is often relatively narrow. For example, the study carried out on serum albumin binding was restricted to compounds having log *P* values in the range 0.78–3.82. If these studies were to be extended to include compounds with very high log *P* values, then we would see a different picture. The graph would be parabolic, as shown in Fig. 18.3. Here, the biological activity increases as log *P* increases until a maximum value is obtained. The value of log *P* at the maximum (log P^0) represents the optimum partition coefficient for biological

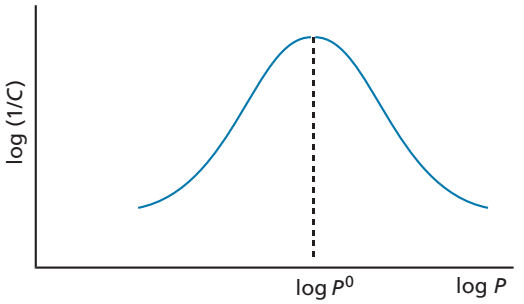


FIGURE 18.3 Parabolic curve of log (1/*C*) vs. log *P*.

activity. Beyond that point, an increase in log *P* results in a decrease in biological activity.

If the partition coefficient is the only factor influencing biological activity, the parabolic curve can be expressed by the equation:

$$\log\left(\frac{1}{C}\right) = -k_1(\log P)^2 + k_2 \log P + k_3$$

Note that the $(\log P)^2$ term has a minus sign in front of it. When *P* is small, the $(\log P)^2$ term is very small and the equation is dominated by the log *P* term. This represents the first part of the graph where activity increases with increasing *P*. When *P* is large, the $(\log P)^2$ term is more significant and eventually 'overwhelms' the log *P* term. This represents the last part of the graph where activity drops with increasing *P*. k_1, k_2, and k_3 are constants and can be determined by a suitable software program.

There are relatively few drugs where activity is related to the log *P* factor alone. Such drugs tend to operate in cell membranes where hydrophobicity is the dominant feature controlling their action. The best example of drugs which operate in cell membranes are the general anaesthetics. Although they also bind to $GABA_A$ receptors, general anaesthetics are thought to function by entering the central nervous system (CNS) and 'dissolving' into cell membranes where they affect membrane structure and nerve function. In such a scenario, there are no specific drug–receptor interactions and the mechanism of the drug is controlled purely by its ability to enter cell membranes (i.e. its hydrophobic character). The general anaesthetic activity of a range of ethers was found to fit the following parabolic equation:

$$\log\left(\frac{1}{C}\right) = -0.22\ (\log P)^2 + 1.04\ \log\ P + 2.16$$

According to this equation, anaesthetic activity increases with increasing hydrophobicity (*P*), as determined by the log *P* factor. The negative $(\log P)^2$ factor shows that the relationship is parabolic and that there is an optimum value for log *P* (log P^0) beyond which increasing hydrophobicity causes a decrease in anaesthetic activity. With this equation, it is now possible to predict the anaesthetic activity of other compounds, given their partition coefficients. However, there are limitations. The equation is derived purely for anaesthetic ethers and is not applicable to other structural types of anaesthetics. This is generally true in QSAR studies. The procedure works best if it is applied to a series of compounds which have the same general structure.

However, QSAR studies *have* been carried out on other structural types of general anaesthetics, and a parabolic curve has been obtained in each case. Although the constants for each equation are different, it is significant that the optimum hydrophobicity (represented by log P^0) for anaesthetic activity is close to 2.3, regardless of the class of anaesthetic being studied. This finding suggests that all general anaesthetics operate in a similar fashion, controlled by the hydrophobicity of the structure.

Because different anaesthetics have similar log P^0 values, the log *P* value of any compound can give some idea of its potential potency as an anaesthetic. For example, the log *P* values of the gaseous anaesthetics **ether**, **chloroform**, and **halothane** are 0.98, 1.97, and 2.3 respectively. Their anaesthetic activity increases in the same order.

As general anaesthetics have a simple mechanism of action based on the efficiency with which they enter the CNS, it implies that log *P* values should give an indication of how easily any compound can enter the CNS. In other words, compounds having a log *P* value close to 2 should be capable of entering the CNS efficiently. This is generally found to be true. For example, the most potent barbiturates for sedative and hypnotic activity are found to have log *P* values close to 2.

As a rule of thumb, drugs which are to be targeted for the CNS should have a log *P* value of approximately 2. Conversely, drugs which are designed to act elsewhere in the body should have log *P* values significantly different from 2 in order to avoid possible CNS side effects (e.g. drowsiness) (see Box 18.1).

18.2.1.2 The substituent hydrophobicity constant (π)

We have seen how the hydrophobicity of a compound can be quantified using the partition coefficient *P*. In order to get *P*, we have to measure it experimentally and that means that we have to synthesize the compounds. It would be much better if we could calculate *P* theoretically and decide in advance whether the compound is worth synthesizing. QSAR would then allow us to target the

BOX 18.1 Altering log *P* to remove central nervous system side effects

The cardiotonic agent (I) was found to produce 'bright visions' in some patients, which implied that it was entering the central nervous system (CNS). This was supported by the fact that the log *P* value of the drug was 2.59. In order to prevent the drug entering the CNS, the 4-OMe group was replaced by a 4-S(O)Me group. This particular group is approximately the same size as the methoxy group, but more hydrophilic. The log *P* value of the new drug (**sulmazole**) was found to be 1.17. The drug was now too hydrophilic to enter the CNS and was free of CNS side effects.

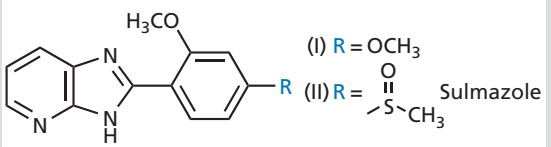


FIGURE 1 Cardiotonic agents.

most promising looking structures. For example, if we were planning to synthesize a range of barbiturate structures, we could calculate log *P* values for them all and concentrate on the structures which had log *P* values closest to the optimum log P^0 value for barbiturates.

Partition coefficients can be calculated by knowing the contribution that various substituents make to hydrophobicity. This contribution is known as the **substituent hydrophobicity constant (π)** and is a measure of how hydrophobic a substituent is relative to hydrogen. The value can be obtained as follows. Partition coefficients are measured experimentally for a standard compound, such as benzene, with and without a substituent (X). The hydrophobicity constant (π_X) for the substituent (X) is then obtained using the following equation:

$$\pi_X = \log P_X - \log P_H$$

where P_H is the partition coefficient for the standard compound and P_X is the partition coefficient for the standard compound with the substituent.

A positive value of π indicates that the substituent is more hydrophobic than hydrogen; a negative value indicates that the substituent is less hydrophobic. The π values for a range of substituents are shown in Table 18.1. These π values are characteristic for the substituent and can be used to calculate how the partition coefficient of a drug would be affected if these substituents were present. The *P* value for the lead compound would have to be measured experimentally, but, once that is known, the *P* value for analogues can be calculated quite simply.

As an example, consider the log *P* values for benzene (log *P* = 2.13), chlorobenzene (log *P* = 2.84), and benzamide (log *P* = 0.64) (Fig. 18.4). Benzene is the parent compound, and the substituent constants for Cl and $CONH_2$ are 0.71 and −1.49 respectively. Having obtained these values, it is now possible to calculate the theoretical log *P* value for *meta*-chlorobenzamide:

$$\begin{aligned}\log P_{(\text{chlorobenzamide})} &= \log P_{(\text{benzene})} + \pi_{Cl} + \pi_{CONH_2} \\ &= 2.13 + 0.17 + (-1.49) \\ &= 1.35\end{aligned}$$

The observed log *P* value for this compound is 1.51.

It should be noted that π values for aromatic substituents are different from those used for aliphatic substituents. Furthermore, neither of these sets of π values are in fact true constants, and they are accurate only for the structures from which they were derived. They can be used as good approximations when studying other structures, but it is possible that the values will have to be adjusted in order to get accurate results.

In order to distinguish calculated log *P* values from experimental ones, the former are referred to as **Clog** *P* values. There are also software programs which will calculate Clog *P* values for a given structure.

Test your understanding and practise your molecular modelling with Exercise 18.2.

18.2.1.3 *P* versus π

QSAR equations relating biological activity to the partition coefficient *P* have already been described, but there is no reason why the substituent hydrophobicity constant π cannot be used in place of *P* if only the substituents are being varied. The equation obtained would be just as relevant as a study of how hydrophobicity affects biological activity. That is not to say that *P* and π are exactly equivalent—different equations would be obtained with different constants. Apart from that, the two factors have different emphases. The partition coefficient *P* is a measure of the drug's overall hydrophobicity and is, therefore, an important measure of how efficiently a drug is transported to its target and bound to its binding site. The π factor measures the hydrophobicity of a specific region on the drug's skeleton and, if it is present in the QSAR equation, it could emphasize important hydrophobic interactions involving that region of the molecule with the binding site.

TABLE 18.1 Values of π for a range of substituents

Group	CH_3	*t*-Bu	OH	OCH_3	CF_3	Cl	Br	F
π (aliphatic substituents)	0.50	1.68	–1.16	0.47	1.07	0.39	0.60	–0.17
π (aromatic substituents)	0.52	1.68	–0.67	–0.02	1.16	0.71	0.86	0.14

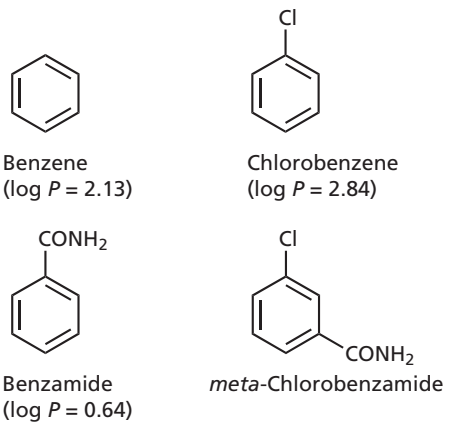


FIGURE 18.4 Values for log *P*.

Most QSAR equations have a contribution from *P* or from π, but there are examples of drugs for which they have only a slight contribution. For example, a study on antimalarial drugs showed very little relationship between antimalarial activity and hydrophobic character. This finding supports the theory that these drugs act in red blood cells, as previous research has shown that the ease with which drugs enter red blood cells is not related to their hydrophobicity.

18.2.2 Electronic effects

The electronic effects of various substituents will clearly have an effect on a drug's ionization or polarity. This, in turn, may have an effect on how easily a drug can pass through cell membranes or how strongly it can interact with a binding site. It is, therefore, useful to measure the electronic effect of a substituent.

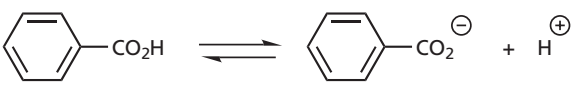


FIGURE 18.5 Ionization of benzoic acid in water.

As far as substituents on an aromatic ring are concerned, the measure used is known as the **Hammett substituent constant** (σ). This is a measure of the electron-withdrawing or electron-donating ability of a substituent, and has been determined by measuring the dissociation of a series of substituted benzoic acids compared with the dissociation of benzoic acid itself.

Benzoic acid is a weak acid and only partially ionizes in water (Fig. 18.5). An equilibrium is set up between the ionized and non-ionized forms, where the relative proportion of these species is known as the **equilibrium** or **dissociation constant** K_H (the subscript H signifies that there are no substituents on the aromatic ring).

$$K_H = \frac{[PhCO_2^-]}{[PhCO_2H]}$$

When a substituent is present on the aromatic ring, this equilibrium is affected. Electron-withdrawing groups, such as a nitro group, result in the aromatic ring having a stronger electron-withdrawing and stabilizing influence on the carboxylate anion, and so the equilibrium will shift more to the ionized form. Therefore, the substituted benzoic acid is a stronger acid and has a larger K_X value (X represents the substituent on the aromatic ring) (Fig. 18.6).

If the substituent X is an electron-donating group such as an alkyl group, then the aromatic ring is less able

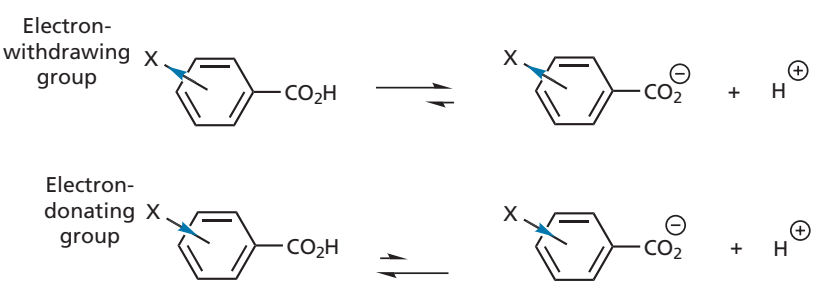


FIGURE 18.6 Position of equilibrium dependent on substituent group X.

to stabilize the carboxylate ion. The equilibrium shifts to the left indicating a weaker acid with a smaller K_X value (Fig. 18.6).

The Hammett substituent constant (σ_X) for a particular substituent (X) is defined by the following equation:

$$\sigma_X = \log \frac{K_X}{K_H} = \log K_X - \log K_H$$

Benzoic acids containing electron-withdrawing substituents will have larger K_X values than benzoic acid itself (K_H) and, therefore, the value of σ_X for an electron-withdrawing substituent will be positive. Substituents such as Cl, CN, or CF_3 have positive σ values.

Benzoic acids containing electron-donating substituents will have smaller K_X values than benzoic acid itself and, hence, the value of σ_X for an electron-donating substituent will be negative. Substituents such as Me, Et, and *t*-Bu have negative values of σ. The Hammett substituent constant for H is zero.

The Hammett substituent constant takes into account both resonance and inductive effects. Therefore, the value of σ for a particular substituent will depend on whether the substituent is *meta* or *para*. This is indicated by the subscript *m* or *p* after the σ symbol. For example, the nitro substituent has $\sigma_p = 0.78$ and $\sigma_m = 0.71$. In the *meta* position, the electron-withdrawing power is due to the inductive influence of the substituent, whereas at the *para* position inductive and resonance both play a part and so the σ_p value is greater (Fig. 18.7).

For the hydroxyl group $\sigma_m = 0.12$ and $\sigma_p = -0.37$. At the *meta* position, the influence is inductive and electron-withdrawing. At the *para* position, the electron-donating influence due to resonance is more significant than the electron-withdrawing influence due to induction (Fig. 18.8).

Most QSAR studies start off by considering σ, and if there is more than one substituent the σ values can be summed (Σσ). However, as more compounds are synthesized, it is possible to fine-tune the QSAR equation. As mentioned above, σ is a measure of a substituent's inductive and resonance electronic effects. With more detailed studies, the inductive and resonance effects can be considered separately. Tables of constants are available which quantify a substituent's inductive effect (*F*) and its resonance effect (*R*). In some cases, it might be found that a substituent's effect on activity is due to *F* rather than *R*, and *vice versa*. It might also be found that a substituent has a more significant effect at a particular position on the ring and this can also be included in the equation.

There are limitations to the electronic constants described so far. For example, Hammett substituent constants cannot be measured for *ortho* substituents as such substituents have an important steric, as well as electronic, effect.

There are very few drugs whose activities are solely influenced by a substituent's electronic effect, as hydrophobicity usually has to be considered as well. Those that do are generally operating by a mechanism whereby they do not have to cross any cell membranes (see Box 18.2). Alternatively, *in vitro* studies on isolated enzymes may result in QSAR equations lacking the hydrophobicity factor, as there are no cell membranes to be considered.

The constants σ, *R*, and *F* can only be used for aromatic substituents and are, therefore, only suitable for drugs containing aromatic rings. However, a series of

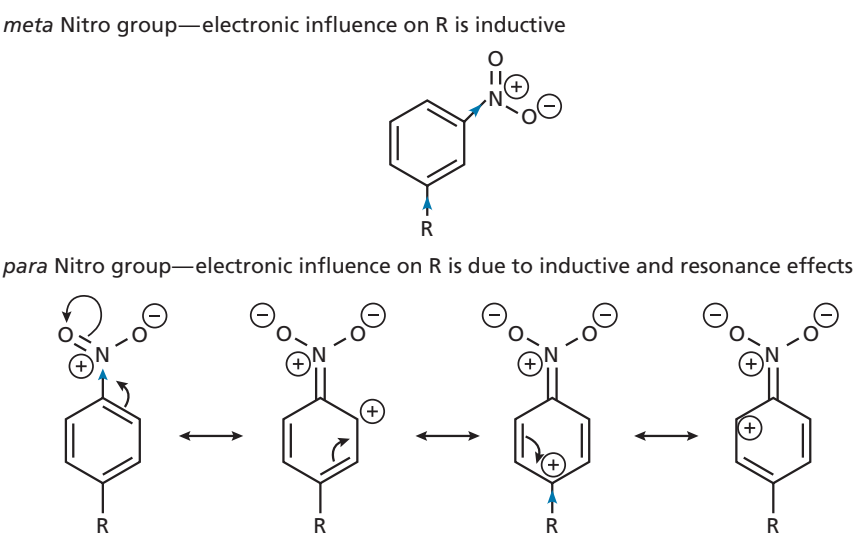
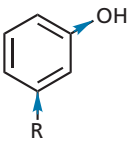


FIGURE 18.7 Substituent effects of a nitro group at the *meta* and *para* positions.

meta Hydroxyl group—electronic influence on R is inductive



para Hydroxyl group—electronic influence on R dominated by resonance effects

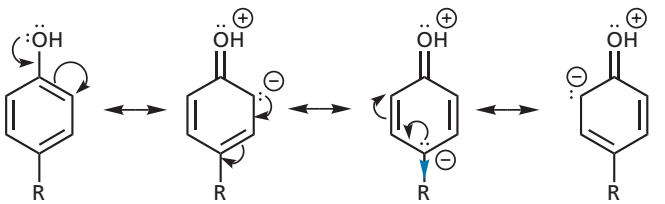


FIGURE 18.8 Substituent effects of a phenol at the *meta* and *para* positions.

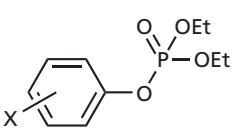
BOX 18.2 Insecticidal activity of diethyl phenyl phosphates

The insecticidal activity of diethyl phenyl phosphates is one of the few examples where activity is related to electronic factors alone:

$$\log\left(\frac{1}{C}\right) = 2.282\sigma - 0.348. \quad (r^2\ 0.952,\ r\ 0.976, s\ 0.286)$$

The equation reveals that substituents with a positive value for σ (i.e. electron-withdrawing groups) will increase activity. The fact that a hydrophobic parameter is not present is a good indication that the drugs do not have to pass into, or through, a cell membrane to have activity. In fact, these drugs are known to act on an enzyme called **acetylcholinesterase** which is situated on the outside of cell membranes (section 22.12).

The value of *r* is close to 1, which demonstrates that the line is a good fit, and the value of r^2 demonstrates that 95% of the data is accounted for by the σ parameter.



Diethyl phenyl phosphates

aliphatic electronic substituent constants are available. These were obtained by measuring the rates of hydrolysis for a series of aliphatic esters (Fig. 18.9). Methyl ethanoate is the parent ester and it is found that the rate of hydrolysis is affected by the substituent X. The extent to which the rate of hydrolysis is affected is a measure of the substituent's electronic effect at the site of reaction (i.e. the ester group). The electronic effect is purely inductive and is given the symbol σ_I. Electron-donating groups reduce the rate of hydrolysis and, therefore, have negative values. For example, σ_I values for methyl, ethyl, and propyl are –0.04, –0.07, and –0.36 respectively. Electron-withdrawing groups increase the rate of hydrolysis and have positive values. The σ_I values for NMe_3^+ and CN are 0.93 and 0.53 respectively.

It should be noted that the inductive effect is not the only factor affecting the rate of hydrolysis. The substituent may also have a steric effect. For example, a bulky substituent may shield the ester from attack and lower the rate of hydrolysis. It is, therefore, necessary to separate out these two effects. This can be done by measuring hydrolysis rates under both basic and acidic conditions. Under basic conditions, steric and electronic factors are important, whereas under acidic conditions only steric factors are important. By comparing the rates, values for the electronic effect (σ_I), and the steric effect (E_S) (see the next sections) can be determined.

Test your understanding and practise your molecular modelling with Exercise 18.3.

18.2.3 Steric factors

The bulk, size, and shape of a drug will influence how easily it can approach and interact with a binding site.

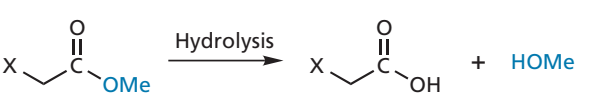


FIGURE 18.9 Hydrolysis of an aliphatic ester.

A bulky substituent may act like a shield and hinder the ideal interaction between a drug and its binding site. Alternatively, a bulky substituent may help to orientate a drug properly for maximum binding and increase activity. Steric properties are more difficult to quantify than hydrophobic or electronic properties. Several methods have been tried, of which three are described here. It is highly unlikely that a drug's biological activity will be affected by steric factors alone, but these factors are frequently found in **Hansch equations** (section 18.3).

18.2.3.1 Taft's steric factor (E_s)

Attempts have been made to quantify the steric features of substituents by using **Taft's steric factor** (E_s). The value for E_s can be obtained by comparing the rates of hydrolysis of substituted aliphatic esters against a standard ester under acidic conditions. Thus,

$$E_S = \log k_X - \log k_O$$

where k_x represents the rate of hydrolysis of an aliphatic ester bearing the substituent X and k_o represents the rate of hydrolysis of the reference ester.

The substituents that can be studied by this method are restricted to those which interact sterically with the tetrahedral transition state of the reaction and not by resonance or internal hydrogen bonding. For example, unsaturated substituents which are conjugated to the ester cannot be measured by this procedure. Examples of E_s values are shown in Table 18.2. Note that the reference ester is X = Me. Substituents such as H and F, which are smaller than a methyl group, result in a faster rate of hydrolysis ($k_x > k_o$), making E_s positive. Substituents which are larger than methyl reduce the rate of hydrolysis ($k_x < k_o$), making E_s negative. A disadvantage of E_s values is that they are a measure of an *intramolecular* steric effect, whereas drugs interact with target binding sites in an *intermolecular* manner. For example, consider the E_s values for *i*-Pr, *n*-Pr, and *n*-Bu. The E_s value for the branched isopropyl group is significantly greater than that for the linear *n*-propyl group since the bulk of the substituent is closer to the reaction centre. Extending the alkyl chain from *n*-propyl to *n*-butyl has little effect on E_s. The larger *n*-butyl group is extended away from the reaction centre and so has little additional steric effect on the rate of hydrolysis. As a result, the E_s value for the *n*-butyl group undervalues the steric effect which this group might have if it was present on a drug approaching a binding site.

18.2.3.2 Molar refractivity

Another measure of the steric factor is provided by a parameter known as **molar refractivity** (***MR***). This is a measure of the volume occupied by an atom or a group of atoms. The *MR* is obtained from the following equation:

$$MR = \frac{(n^2 - 1)}{(n^2 + 2)} \times \frac{MW}{d}$$

where *n* is the index of refraction, *MW* is the molecular weight, and *d* is the density. The term *MW*/*d* defines a volume and the $(n^2 - 1)/(n^2 + 2)$ term provides a correction factor by defining how easily the substituent can be polarized. This is particularly significant if the substituent has π electrons or lone pairs of electrons.

18.2.3.3 Verloop steric parameter

Another approach to measuring the steric factor involves a computer program called **Sterimol**, which calculates steric substituent values (**Verloop steric parameters**) from standard bond angles, van der Waals radii, bond lengths, and possible conformations for the substituent. Unlike E_s, the Verloop steric parameters can be measured for any substituent. For example, the Verloop steric parameters for a carboxylic acid group are demonstrated in Fig. 18.10. L is the length of the substituent and B_1–B_4 are the radii of the group in different dimensions.

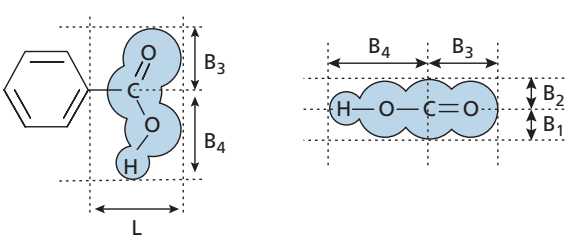


FIGURE 18.10 Verloop parameters for a carboxylic acid group.

TABLE 18.2 Values of E_s for various substituents

Substituent	H	F	Me	Et	n-Pr	n-Bu	i-Pr	i-Bu	Cyclopentyl
E_s	1.24	0.78	0	–0.07	–0.36	–0.39	–0.47	–0.93	–0.51

18.2.4 Other physicochemical parameters

The physicochemical properties most commonly studied by the QSAR approach have been described above, but other properties have been studied including dipole moments, hydrogen bonding, conformations, and interatomic distances. Difficulties in quantifying these properties limit the use of these parameters, however. Several QSAR formulae have been developed based on the highest occupied and/or the lowest unoccupied molecular orbitals of the test compounds. The calculation of these orbitals can be carried out using semi-empirical quantum mechanical methods (section 17.7.3). **Indicator variables** for different substituents can also be used. These are described in section 18.7.

18.3 Hansch equation

In section 18.2, we looked at the physicochemical properties commonly used in QSAR studies and how it is possible to quantify them. In a situation where biological activity is related to only one such property, a simple equation can be drawn up. The biological activity of most drugs, however, is related to a combination of physicochemical properties. In such cases, simple equations involving only one parameter are relevant only if the other parameters are kept constant. In reality, this is not easy to achieve and equations which relate biological activity to a number of different parameters are more common (Box 18.3). These equations are known as Hansch equations and they usually relate biological activity to the most commonly used physicochemical properties (log *P*, π, σ, and a steric factor). If the range of hydrophobicity values is limited to a small range then the equation will be linear, as follows:

$$\log\left(\frac{1}{C}\right) = k_1 \log P + k_2\sigma + k_3 E_s + k_4$$

If the log *P* values are spread over a large range, then the equation will be parabolic for the same reasons described in section 18.2.1,

$$\log\left(\frac{1}{C}\right) = -k_1(\log P)^2 + k_2 \log P + k_3\sigma + k_4 E_s + k_5$$

The constants k_1–k_5 are determined by computer software in order to get the best-fitting equation. Not all the parameters will necessarily be significant. For example, the adrenergic blocking activity of β-halo-arylamines (Fig. 18.11) was related to π and σ and did not include a steric factor. This equation tells us that biological activity increases if the substituents have a positive π value and a negative σ value. In other words, the substituents should be hydrophobic and electron donating.

When carrying out a Hansch analysis, it is important to choose the substituents carefully to ensure that the change in biological activity can be attributed to a particular parameter. There are plenty of traps for the unwary. Take, for example, drugs which contain an amine group. One of the studies most frequently carried out on amines is to synthesize analogues containing a homologous series of alkyl substituents on the nitrogen atom (i.e. Me, Et, *n*-Pr, *n*-Bu). If activity increases with the chain length of the substituent, is it due to increasing hydrophobicity, increasing size, or both? If we look at the π and *MR* values of these substituents, we find that both sets of values increase in a similar fashion across the series and we would not be able to distinguish between them (Table 18.3). In this example, a series of substituents would have to be chosen where π and *MR* are not correlated. The substituents H, Me, OMe, $NHCOCH_2$, I, and CN would be more suitable.

Test your understanding and practise your molecular modelling with Exercise 18.4. You might also find exercises 25.4–25.6 and 25.4 useful at this point.

18.4 The Craig plot

Although tables of π and σ factors are readily available for a large range of substituents, it is often easier to visualize the relative properties of different substituents by considering a plot where the y-axis is the value of the σ factor and the x-axis is the value of the π factor. Such a plot is known as a **Craig plot**. The example shown in Fig. 18.12 is the Craig plot for the σ and π factors of *para*-aromatic substituents. There are several advantages to the use of such a Craig plot.

- The plot shows clearly that there is no overall relationship between π and σ. The various substituents are scattered around all four quadrants of the plot.
- It is possible to tell at a glance which substituents have positive π and σ parameters, which substituents have negative π and σ parameters, and which substituents have one positive and one negative parameter.
- It is easy to see which substituents have similar π values. For example, the ethyl, bromo, trifluoromethyl, and trifluoromethylsulfonyl groups are all approximately on the same vertical line on the plot. In theory, these groups could be interchangeable on drugs where the principal factor affecting biological activity is the π factor. Similarly, groups which form a horizontal line can be identified as being isoelectronic or having similar σ values (e.g. CO_2H, Cl, Br, I).

BOX 18.3 Hansch equation for a series of antimalarial compounds

A series of 102 phenanthrene aminocarbinols was tested for antimalarial activity. In the structure shown, X represents up to four substituents on the left-hand ring while Y represents up to four substituents on the right-hand ring. Experimental log *P* values for the structures were not available and equations were derived which compared the activity with some or all of the following terms:

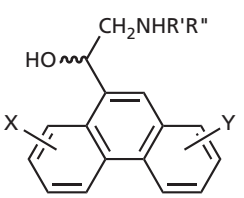
- π_{sum} the π constants for *all* the substituents in the molecule (i.e. all the X and Y substituents, as well as the amino substituents R and R"). This term was used in place of log *P* to represent the overall hydrophobicity for the molecule;
- σ_{sum} the σ constants for *all* the substituents in the molecule;
- $\Sigma\pi_X$ the sum of the π constants for all the substituents X in the left-hand ring;
- $\Sigma\pi_Y$ the sum of the π constants for all the substituents Y in the right-hand ring;
- $\Sigma\pi_{X+Y}$ the sum of the π constants for all the substituents X and Y in both the left and right-hand rings;
- $\Sigma\sigma_{X+Y}$ the sum of the σ constants for all the substituents X and Y in both the left and right-hand rings;
- $\Sigma\sigma_X$ the sum of the σ constants for all the substituents X in the left-hand-ring;
- $\Sigma\sigma_Y$ the sum of the σ constants for all the substituents Y in the right-hand ring.

Equations such as equations 1–3 were derived which matched activity against one of the above terms, but none of them had an acceptable value of r^2.

A variety of other equations were derived which included two of the above terms but these were not satisfactory either. Finally, an equation was derived which contained six terms and proved satisfactory:

$$\log\left(\frac{1}{C}\right) = -0.015(\pi_{sum})^2 + 0.14\pi_{sum} + 0.27\Sigma\pi_X + 0.40\Sigma\pi_Y + 0.65\Sigma\sigma_X + 0.88\Sigma\sigma_Y + 2.34$$
$$(n = 102,\ r = .0913,\ r^2 = 0.834,\ s = 0.258)$$

The equation shows that antimalarial activity increases very slightly as the overall hydrophobicity of the molecule (π_{sum}) increases (the constant 0.14 is low). The $(\pi_{sum})^2$ term shows that there is an optimum overall hydrophobicity for activity and this is found to be 4.44. Activity increases if hydrophobic substituents are present on ring X and, in particular, on ring Y. This could be taken to imply that some form of hydrophobic interaction is involved near both rings. Electron-withdrawing substituents on both rings are also beneficial to activity, more so on ring Y than ring X. The r^2 value is 0.834 which is above the minimum acceptable value of 0.8.

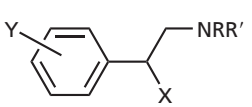


1) $\log\left(\frac{1}{C}\right) = 0.557\Sigma\pi_{x+y} + 2.699$ ($n = 102$, $r = 0.768$, $r^2 = 0.590$, $s = 0.395$)

2) $\log\left(\frac{1}{C}\right) = 0.017\pi_{sum} + 3.324$ ($n = 102$, $r = 0.069$, $r^2 = 0.005$, $s = 0.616$)

3) $\log\left(\frac{1}{C}\right) = 1.218\sigma_{sum} + 2.721$ ($n = 102$, $r = 0.814$, $r^2 = 0.663$, $s = 0.359$)

FIGURE 1 Phenanthrene aminocarbinols.



β-Halo-arylamines

$$\log\left(\frac{1}{C}\right) = 1.22\pi - 1.59\sigma + 7.89$$
$$(n = 22,\ r^2 = 0.841,\ s = 0.238)$$

FIGURE 18.11 QSAR equation for β-halo-arylamines.

- The Craig plot is useful in planning which substituents should be used in a QSAR study. In order to derive the most accurate equation involving π and σ, analogues should be synthesized with substituents from each quadrant. For example, halogen substituents are useful representatives of substituents with increased

TABLE 18.3 Values for π and MR for a series of substituents

Substituent	H	Me	Et	n-Pr	n-Bu	OMe	$NHCONH_2$	I	CN
π	0.00	0.56	1.02	1.50	2.13	–0.02	–1.30	1.12	–0.57
MR	0.10	0.56	1.03	1.55	1.96	0.79	1.37	1.39	0.63

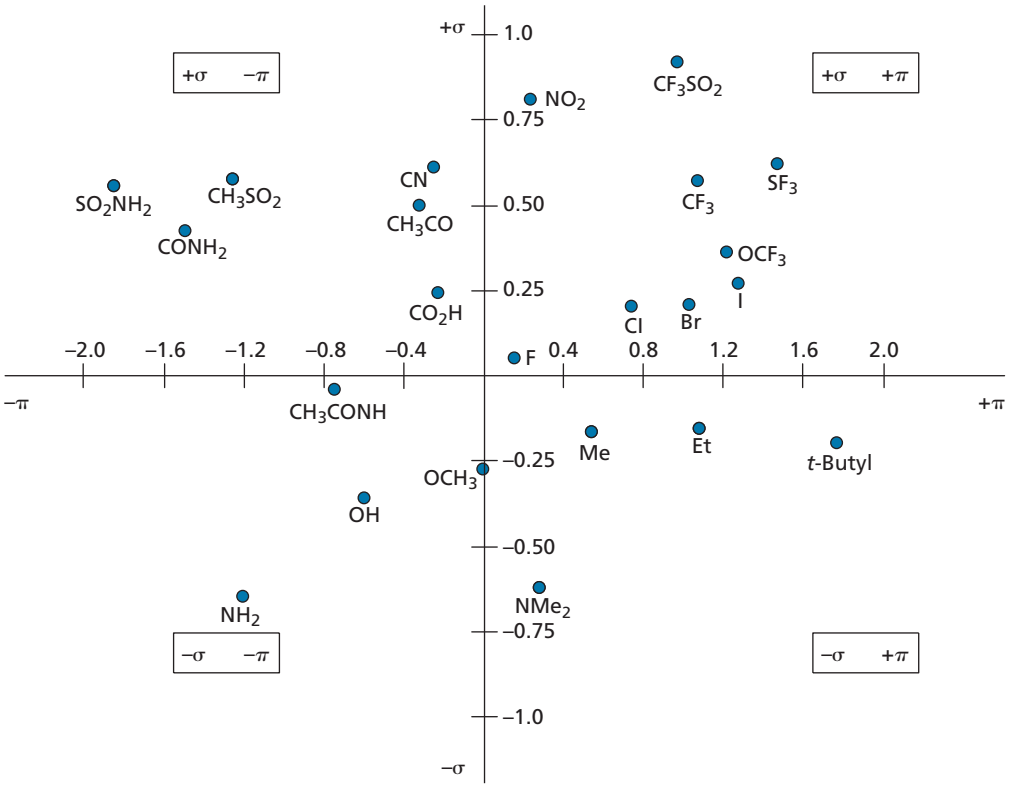


FIGURE 18.12 Craig plot comparing the values of σ and π for various substituents.

hydrophobicity and electron-withdrawing properties (positive π and positive σ), whereas an OH substituent has more hydrophilic and electron-donating properties (negative π and negative σ). Alkyl groups are examples of substituents with positive π and negative σ values, whereas acyl groups have negative π and positive σ values.

- Once the Hansch equation has been derived, it will show whether π or σ should be negative or positive in order to get good biological activity. Further developments would then concentrate on substituents from the relevant quadrant. For example, if the equation shows that positive π and positive σ values are necessary, then further substituents should only be taken from the top right quadrant.

Craig plots can also be drawn up to compare other sets of physicochemical parameters, such as hydrophobicity and *MR*.

18.5 The Topliss scheme

In certain situations, it might not be feasible to make the large range of structures required for a Hansch equation. For example, the synthetic route involved might be so difficult that only a few structures can be made in a limited time. In these circumstances, it would be useful to test compounds for biological activity as they are synthesized and to use these results to determine the next analogue to be synthesized.

A **Topliss scheme** is a 'flow diagram' which allows such a procedure to be followed. There are two Topliss schemes, one for aromatic substituents (Fig. 18.13) and one for aliphatic side-chain substituents (Fig. 18.14). The schemes were drawn up by considering the hydrophobicity and electronic factors of various substituents, and are designed such that the optimum substituent can be found as efficiently as possible. They are not meant to be a replacement for a full Hansch analysis, however. Such an analysis would be carried out in due course, once a suitable number of structures have been synthesized.

The Topliss scheme for aromatic substituents (Fig. 18.13) assumes that the lead compound has been tested for biological activity and contains a monosubstituted aromatic ring. The first analogue in the scheme is the 4-chloro derivative, as this derivative is usually easy to synthesize. The chloro substituent is more hydrophobic and electron-withdrawing than hydrogen and, therefore, π and σ are positive.

Once the chloro analogue has been synthesized, the biological activity is measured. There are three

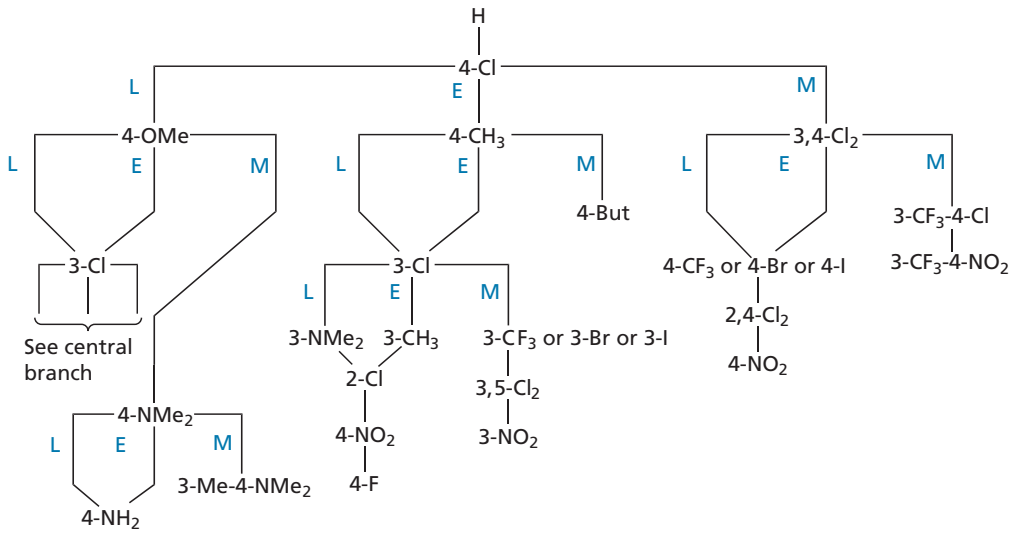


FIGURE 18.13 Topliss scheme for aromatic substituents.

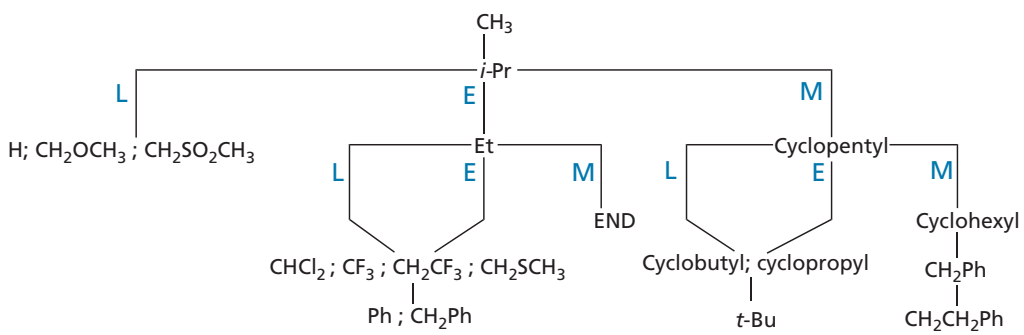


FIGURE 18.14 Topliss scheme for aliphatic side-chain substituents.

possibilities. The analogue will have less activity (L), equal activity (E), or more activity (M). The type of activity observed will determine which branch of the Topliss scheme is followed next.

If the biological activity increases, the (M) branch is followed and the next analogue to be synthesized is the 3,4-dichloro-substituted analogue. If, however, the activity stays the same, then the (E) branch is followed and the 4-methyl analogue is synthesized. Finally, if activity drops, the (L) branch is followed and the next analogue is the 4-methoxy analogue. Biological results from the second analogue now determine the next branch to be followed in the scheme.

What is the rationale behind this?

Let us consider the situation where the 4-chloro derivative increases in biological activity. The chloro substituent has positive π and σ values, which implies that one, or both, of these properties are important to biological activity. If both are important, then adding a second chloro group should increase biological activity yet further. If it does, substituents are varied to increase the π and σ values even further. If it does not, then an unfavourable steric interaction or excessive hydrophobicity is indicated. Further modifications then test the relative importance of π and steric factors.

Now consider the situation where the 4-chloro analogue drops in activity. This suggests either that negative π and/or σ values are important to activity or that a *para* substituent is sterically unfavourable. It is assumed that an unfavourable σ effect is the most likely reason for the reduced activity and so the next substituent is one with a negative σ factor (i.e. 4-OMe). If activity improves, further changes are suggested to test the relative importance of the σ and π factors. However, if the 4-OMe group does not improve activity, it is assumed that an unfavourable steric factor is at work and the next substituent is a 3-chloro group. Modifications of this group would then be carried out in the same way as shown in the centre branch of Fig. 18.13.

The last scenario is where the activity of the 4-chloro analogue is little changed from the lead compound. This could arise from the drug requiring a positive π value and a negative σ value. As both values for the chloro group are positive, the beneficial effect of the positive π value might be cancelled out by the detrimental effects of a positive σ value. The next substituent to try in that case is the 4-methyl group which has the necessary positive π value and negative σ value. If this still has no beneficial effect, then it is assumed that there is an unfavourable steric interaction at the *para* position and the 3-chloro substituent is chosen next. Further changes continue to vary the relative values of the π and σ factors.

The validity of the Topliss scheme was tested by looking at structure–activity results for various drugs which had been reported in the literature. For example, the biological activities of 19 substituted benzenesulphonamides (Fig. 18.15) have been reported. The second most active compound was the nitro-substituted analogue, which would have been the fifth compound synthesized if the Topliss scheme had been followed.

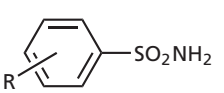
Another example comes from the anti-inflammatory activities of substituted aryltetrazolylalkanoic acids (Fig. 18.16), of which 28 were synthesized. Using the Topliss scheme, three out of the four most active structures would have been prepared from the first eight compounds synthesized.

The Topliss scheme for aliphatic side chains (Fig. 18.14) was set up following a similar rationale to the aromatic scheme and is used in the same way for side groups attached to a carbonyl, amino, amide, or similar functional group. The scheme attempts to differentiate only between the hydrophobic and electronic effects of substituents, and not their steric properties. Thus, the substituents involved have been chosen to try to minimize any steric differences. It is assumed that the lead compound has a methyl group. The first analogue suggested is the isopropyl analogue. This has an increased π value and, in most cases, would be expected to increase activity. It has been found from experience that the hydrophobicity of most lead compounds is less than optimum.

Let us concentrate first of all on the situation where activity increases. Following this branch, a cyclopentyl group is now used. A cyclic structure is used as it has a larger π value, but keeps any increase in steric factor to a minimum. If activity rises again, more hydrophobic substituents are tried. If activity does not rise, then there could be two explanations. Either the optimum hydrophobicity has been passed or there is an electronic effect (σ_I) at work. Further substituents are then used to determine which is the correct explanation.

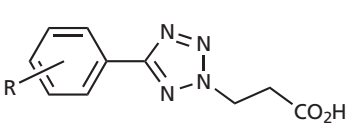
Let us now look at the situation where the activity of the isopropyl analogue stays much the same. The most likely explanation is that the methyl group and the isopropyl group are on either side of the hydrophobic optimum. Therefore, an ethyl group is used next, as it has an intermediate π value. If this does not lead to an improvement, it is possible that there is an unfavourable electronic effect. The groups used have been electron-donating and so electron-withdrawing groups with similar π values are now suggested.

Finally, we shall look at the case where activity drops for the isopropyl group. In this case, hydrophobic and/or electron-donating groups could be bad for activity



Order of synthesis	R	Biological activity	High potency
1	H	–	
2	4-Cl	More	
3	3,4-Cl_2	Less	
4	4-Br	Equal	
5	4-NO_2	More	*

FIGURE 18.15 The order of benzenesulphonamide synthesis, as directed by the Topliss scheme.



Order of synthesis	R	Biological activity	High potency
1	H	–	
2	4-Cl	Less	
3	4-OMe	Less	
4	3-Cl	More	*
5	3-CF_3	Less	
6	3-Br	More	*
7	3-I	Less	
8	3,5-Cl_2	More	*

FIGURE 18.16 The order of synthesis for substituted aryltetrazolylakonoic acids, as directed by the Topliss scheme.

and the groups suggested are suitable choices for further development.

18.6 Bioisosteres

Tables of substituent constants are available for various physicochemical properties. A knowledge of these constants allows the medicinal chemist to identify substituents which may be potential bioisosteres. Thus, the substituents CN, NO_2, and COMe have similar hydrophobic, electronic, and steric factors, and might be interchangeable. Such interchangeability was observed in the development of **cimetidine** and cimetidine analogues (sections 25.2.6 and 25.2.8). The important thing to note is that groups can be bioisosteric in some situations, but not others. Consider, for example, the table shown in Fig. 18.17.

This table shows physicochemical parameters for six different substituents. If the most important physicochemical parameter for biological activity is σ_p, then the $COCH_3$ group (0.50) would be a reasonable bioisostere for the $SOCH_3$ group (0.49). If, however, the dominant parameter is π, then a more suitable bioisostere for $SOCH_3$ (−1.58) would be SO_2CH_3 (−1.63).

18.7 The Free-Wilson approach

In the **Free-Wilson approach** to QSAR, the biological activity of a parent structure is measured then compared with the activities of a range of substituted analogues. An equation is then derived which relates biological activity to the presence, or otherwise, of particular substituents (X_1–X_n).

$$\text{Activity} = k_1X_1 + k_2X_2 + K_3X_3 + \ldots\ldots\ldots\ldots + k_nX_n + Z.$$

In this equation, X_n is defined as an **indicator variable** and is given the value 1 or 0, depending on whether the substituent (n) is present or not. The contribution that each substituent makes to the activity is determined by the value of k_n. Z is a constant representing the average activity of the structures studied.

Since the approach considers the overall effect of a substituent to biological activity rather than its various physicochemical properties, there is no need for physicochemical constants and tables, and the method only requires experimental measurements of biological activity. This is particularly useful when trying to quantify the effect of unusual substituents that are not listed in the tables, or when quantifying specific molecular features which cannot be tabulated.

The disadvantage in the approach is the large number of analogues which have to be synthesized and tested to make the equation meaningful. For example, each of the terms k_nX_n refers to a specific substituent at a specific position in the parent structure. Therefore, analogues would not only have to have different substituents, but also have them at different positions of the skeleton.

Another disadvantage is the difficulty in rationalizing the results and explaining why a substituent at a particular position is good or bad for activity. Finally, the effects of different substituents may not be additive. There may be intramolecular interactions which affect activity.

Nevertheless, indicator variables can be useful in certain situations and they can also be used as part of a Hansch equation. An example of this can be seen in the later case study (section 18.9).

18.8 Planning a QSAR study

When starting a QSAR study it is important to decide which physicochemical parameters are going to be studied and to plan the analogues such that the parameters under study are suitably varied. For example, it would be pointless to synthesize analogues where the hydrophobicity and steric volume of the substituents are correlated if these two parameters are to go into the equation.

It is also important to make enough structures to make the results statistically meaningful. As a rule of thumb, five structures should be made for every parameter studied. Typically, the initial QSAR study would involve the two parameters π and σ, and, possibly, E_S. Craig plots could be used in order to choose suitable substituents.

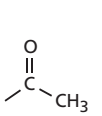
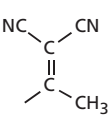
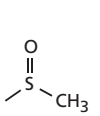
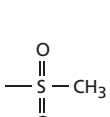
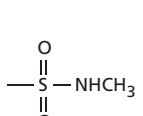
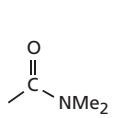
Substituent						
π	−0.55	0.40	−1.58	−1.63	−1.82	−1.51
σ_p	0.50	0.84	0.49	0.72	0.57	0.36
σ_m	0.38	0.66	0.52	0.60	0.46	0.35
MR	11.2	21.5	13.7	13.5	16.9	19.2

FIGURE 18.17 Physicochemical parameters for six substituents.

Certain substituents are worth avoiding in the initial study, as they may have properties other than those being studied. For example, it is best to avoid substituents that might ionize (CO_2H, NH_2, SO_3H) and groups that might be easily metabolized (e.g. esters or nitro groups).

If there are two or more substituents, then the initial equation usually considers the total π and σ contribution.

As more analogues are made, it is often possible to consider the hydrophobic and electronic effect of substituents at specific positions of the molecule. Furthermore, the electronic parameter σ can be split into its inductive and resonance components (*F* and *R*). Such detailed equations may show up a particular localized requirement for activity. For example, a hydrophobic substituent may be favoured in one part of the skeleton, while an electron-withdrawing substituent is favoured at another. In turn, this gives clues about the binding interactions involved between drug and receptor.

18.9 Case study

An example of how a QSAR equation can become more specific as a study develops is demonstrated from work carried out on the anti-allergic activity of a series of pyranenamines (Fig. 18.18). In this study, substituents were varied on the aromatic ring, while the remainder of the molecule was kept constant. Nineteen compounds were synthesized and the first QSAR equation was obtained by considering π and σ:

$$\log\left(\frac{1}{C}\right) = -0.14\ \Sigma\pi - 1.35\ (\Sigma\sigma)^2 - 0.72$$

$$(n\,19, r^2\,0.48, s\,0.47, F_{2,16}\,7.3)$$

where Σπ and Σσ are the total π and σ values for all the substituents present.

The negative coefficient for the π term shows that activity is inversely proportional to hydrophobicity, which is quite unusual. The $(\Sigma\sigma)^2$ term is also quite unusual. It was chosen because there was no simple relationship between activity and σ. In fact, it was observed that activity decreased if the substituent was electron-withdrawing *or* electron-donating. Activity was best with neutral substituents. To take account of this, the $(\Sigma\sigma)^2$ term was introduced. As the coefficient in the equation is negative, activity is lowered if σ is anything other than zero.

A further range of compounds was synthesized with hydrophilic substituents to test this equation, making a total of 61 structures. This resulted in the following inconsistencies.

- The activities for the substituents 3-NHCOMe, 3-NHCOEt, and 3-NHCOPr were all similar, but, according to the equation, these activities should have dropped as the alkyl group got larger as a result of increasing hydrophobicity.
- Activity was greater than expected if there was a substituent such as OH, SH, NH_2, or NHCOR at position 3, 4, or 5.
- The substituent $NHSO_2R$ was bad for activity.
- The substituents 3,5-$(CF_3)_2$ and 3,5-$(NHCOMe)_2$ had much greater activity than expected.
- An acyloxy group at the 4-position resulted in an activity five times greater than predicted by the equation.

These results implied that the initial equation was too simple and that properties other than π and σ were important to activity. At this stage, the following theories were proposed to explain the above results.

- The similar activities for 3-NHCOMe, 3-NHCOEt, and 3-NHCOPr could be due to a steric factor. The substituents had increasing hydrophobicity, which is bad for activity, but they were also increasing in size and it was proposed that this was good for activity. The most likely explanation is that the size of the substituent is forcing the drug into the correct orientation for optimum receptor interaction.
- The substituents which increased activity unexpectedly when they were at positions 3, 4, or 5 are all capable of hydrogen bonding. This suggests an important hydrogen bonding interaction with the receptor. For some reason, the $NHSO_2R$ group is an exception, which implies there is some other unfavourable steric or electronic factor peculiar to this group.
- The increased activity for 4-acyloxy groups was explained by suggesting that these analogues are acting as prodrugs. The acyloxy group is less polar than the hydroxyl group and so these analogues would be expected to cross cell membranes and reach the receptor more efficiently than analogues bearing a free hydroxyl group. At the receptor, the ester group could be hydrolysed to reveal the hydroxyl group which would then take part in hydrogen bonding with the receptor.
- The structures having substituents 3,5-$(CF_3)_2$ and 3,5-$(NHCOMe)_2$ are the only disubstituted structures where a substituent at position 5 has an

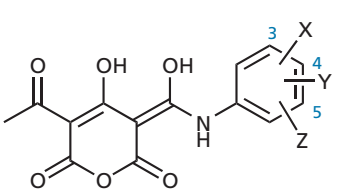


FIGURE 18.18 Structure of pyranenamines.

electron-withdrawing effect, so this feature was also introduced into the next equation.

The revised QSAR equation was as follows:

$$\log\left(\frac{1}{C}\right) = -0.30\sum\pi - 1.5(\sum\sigma)^2 + 2.0(F\text{-}5) + 0.39(345\text{-}HBD) - 0.63(NHSO_2) + 0.78(M\text{-}V) + 0.72(4\text{-}OCO) - 075$$
$$(n\ 61, r^2 0.77, s\ 0.40, F_{7,53}\ 25.1)$$

The π and σ parameters are still present, but a number of new parameters have now been introduced.

- The *F*-5 term represents the inductive effect of a substituent at position 5. The coefficient is positive and large, showing that an electron-withdrawing group increases activity substantially. However, only two compounds of the 61 synthesized had a 5-substituent, so there might be quite an error in this result.
- The *M-V* term represents the volume of any *meta* substituent. The coefficient is positive, indicating that substituents with a large volume at the *meta* position increase activity.
- The advantage of having hydrogen bonding substituents at positions 3, 4, or 5 is accounted for by including a hydrogen bonding term (*345-HBD*). The value of this term depends on the number of hydrogen bonding substituents present. If one such group is present, the *345-HBD* term is 1. If two such groups are present, the parameter is 2. Therefore, for each hydrogen bonding substituent present at positions 3, 4, or 5, log (1/*C*) increases by 0.39. This sort of term is known as an **indicator variable**, which is the basis of the **Free-Wilson** approach described earlier. There is no tabulated value one can use for a hydrogen bonding substituent and so the contribution that this term makes to the biological activity is determined by the value of *k*, and whether the relevant group is present or not. Indicator variables were also used for the following terms.
- The $NHSO_2$ term was introduced because this group was bad for activity despite being capable of hydrogen bonding. The negative coefficient indicates the drop in activity. A figure of 1 is used for any $NHSO_2R$ substituent present, resulting in a drop of activity by 0.63.
- The *4-OCO* term is 1 if an acyloxy group is present at position 4, and so log (1/*C*) is increased by 0.72 if this is the case.

A further 37 structures were synthesized to test steric and *F*-5 parameters, as well as exploring further groups capable of hydrogen bonding. Since hydrophilic substituents were good for activity, a range of very hydrophilic substituents were also tested to see if there was an optimum value for hydrophilicity. The results obtained highlighted one more anomaly, in that two hydrogen bonding groups *ortho* to each other were bad for activity. This was attributed to the groups hydrogen bonding with each other rather than to the receptor. A revised equation was obtained as follows:

$$\log\left(\frac{1}{C}\right) = -0.034(\sum\pi)^2 - 0.33(\sum\pi) + 4.3(F\text{-}5) - 1.3(R\text{-}5) - 1.7(\sum\sigma)^2 + 0.73(345\text{-}HBD) - 0.86(HB\text{-}INTRA) - 0.69(NHSO_2) + 0.72(4\text{-}OCO) - 0.59$$
$$(n\ 98, r^2 0.75, s\ 0.48, F_{9,88}\ 28.7)$$

The main points of interest from this equation are as follows.

- Increasing the hydrophilicity of substituents allowed the identification of an optimum value for hydrophobicity ($\Sigma\pi = -5$) and introduced the $(\Sigma\pi)^2$ parameter into the equation. The value of −5 is remarkably low and indicates that the receptor site is hydrophilic.
- As far as electronic effects are concerned, it is revealed that the resonance effects of substituents at the 5-position also have an influence on activity.
- The unfavourable situation where two hydrogen bonding groups are *ortho* to each other is represented by the *HB-INTRA* parameter. This parameter is given the value 1 if such an interaction is possible, and the negative constant (−0.86) shows that such interactions decrease activity.
- It is interesting to note that the steric parameter is no longer significant and has disappeared from the equation.

The compound having the greatest activity has two $NHCOCH(OH)CH_2OH$ substituents at the 3- and 5-positions and is 1000 times more active than the original lead compound. The substituents are very polar and are not ones that would be used normally. They satisfy all the requirements determined by the QSAR study. They are highly polar groups which can take part in hydrogen bonding. They are *meta* with respect to each other, rather than *ortho*, to avoid undesirable intramolecular hydrogen bonding. One of the groups is at the 5-position and has a favourable *F*-5 parameter. Together, the two groups have a negligible $(\Sigma\sigma)^2$ value. Such an analogue would certainly not have been obtained by trial and error, and this example demonstrates the strengths of the QSAR approach.

All the evidence from this study suggests that the aromatic ring of this series of compounds fits into a hydrophilic pocket in the receptor which contains polar groups capable of hydrogen bonding.

It is further proposed that a positively charged residue such as arginine, lysine, or histidine might be present in the pocket which could interact with an electronegative substituent at position 5 of the aromatic ring (Fig. 18.19).

This example demonstrates that QSAR studies and computers are powerful tools in medicinal chemistry. However, it also shows that the QSAR approach is a long way from replacing the human factor. One cannot put a series of facts and figures into a computer and expect it to magically produce an instant explanation of how a drug works. The medicinal chemist still has to interpret results, propose theories, and test those theories by incorporating the correct parameters into the QSAR equation. Imagination and experience still count for a great deal.

KEY POINTS

- QSAR relates the physicochemical properties of a series of drugs to their biological activity by means of a mathematical equation.
- The commonly studied physicochemical properties are hydrophobicity, electronic factors, and steric factors.
- The partition coefficient is a measure of a drug's overall hydrophobicity. Values of log *P* are used in QSAR equations, with larger values indicating greater hydrophobicity.
- The substituent hydrophobicity constant is a measure of the hydrophobic character of individual substituents. The value is different for aliphatic and aromatic substituents, and is only directly relevant to the class of structures from which the values were derived. Positive values represent substituents more hydrophobic than hydrogen; negative values represent substituents more hydrophilic than hydrogen.
- The Hammett substituent constant is a measure of how electron-withdrawing or electron-donating an aromatic substituent is. It is measured experimentally and is dependent on the relative position of the substituent on the ring. The value takes into account both inductive and resonance effects.
- The parameters *F* and *R* are constants quantifying the inductive and resonance effects of an aromatic substituent.
- The inductive effect of aliphatic substituents can be measured experimentally and tabulated.
- Steric factors can be measured experimentally or calculated using physical parameters or computer software.
- The Hansch equation is a mathematical equation which relates a variety of physicochemical parameters to biological activity for a series of related structures.
- The Craig plot is a visual comparison of two physicochemical properties for a variety of substituents. It facilitates the choice of substituents for a QSAR study such that the values of each property are not correlated.
- The Topliss scheme is used when structures can only be synthesized and tested one at a time. The scheme is a guide to which analogue should be synthesized next in order to get good activity. There are different schemes for aromatic and aliphatic substituents.
- Indicator variables are used when there are no tabulated or experimental values for a particular property or substituent. The Free Wilson approach to QSAR only uses indicator variables, whereas the Hansch approach can use a mixture of indicator variables and physicochemical parameters.

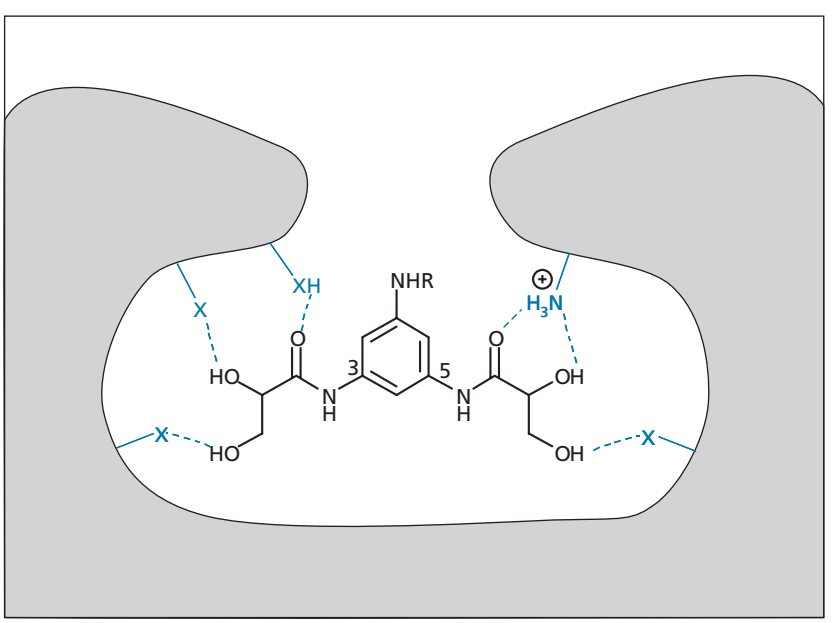


FIGURE 18.19 Hypothetical binding interactions between a pyranenamine and the target binding site.

18.10 Three-dimensional QSAR

In recent years, a method known as **three-dimensional (3D) QSAR** has been developed in which the 3D properties of a molecule are considered as a whole rather than considering individual substituents or moieties. This has proved remarkably useful in the design of new drugs. Moreover, the necessary software and hardware are readily affordable and relatively easy to use. The philosophy of 3D QSAR revolves around the assumption that the most important features about a molecule are its overall size and shape, and its electronic properties.

If these features can be defined, then it is possible to study how they affect biological properties. There are several approaches to 3D QSAR, but the method which has gained ascendancy was developed by the company Tripos and is known as **comparative molecular field analysis** (**CoMFA**). CoMFA methodology is based on the assumption that drug–receptor interactions are non-covalent and that changes in biological activity correlate with the changes in the steric and/or electrostatic fields of the drug molecules.

18.10.1 Defining steric and electrostatic fields

The steric and electrostatic fields surrounding a molecule can be measured and defined using the grid and probe method described in section 17.7.5. This can be repeated for all the molecules in the 3D QSAR study, but it is crucial that the molecules are all in their **active conformation**, and that they are all positioned within the grid in exactly the same way. In other words, they must all be correctly aligned. Identifying a **pharmacophore** (section 13.2) that is common to all the molecules can assist in this process (Fig. 18.20).

The pharmacophore is placed into the grid and its position is kept constant such that it acts as a reference point when positioning each molecule into the lattice. For each molecule studied, the active conformation and pharmacophore is identified and then the molecule is placed into the lattice such that its pharmacophore matches the reference pharmacophore (Fig. 18.21). Once a molecule has been placed into the lattice, the steric and electrostatic fields around it are measured as described in section 17.7.5.

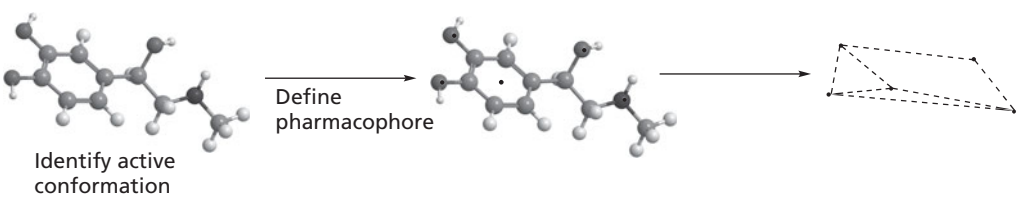


FIGURE 18.20 Identification of the active conformation and pharmacophore.

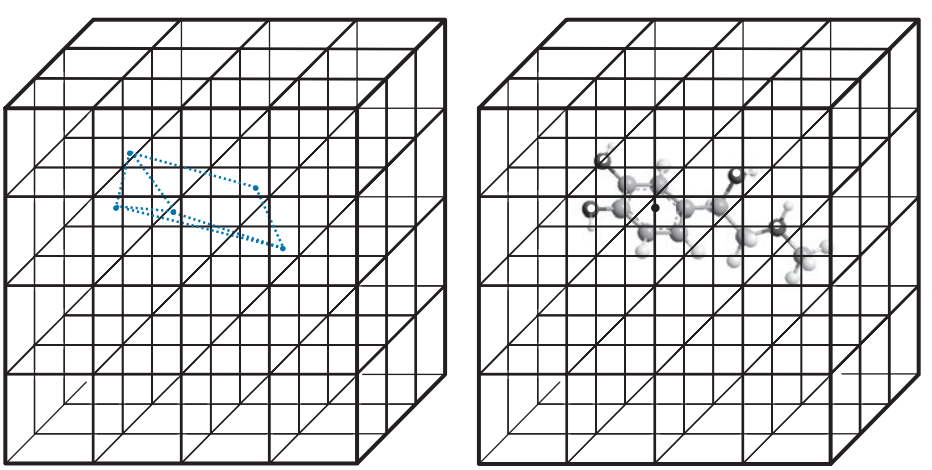


FIGURE 18.21 Positioning a pharmacophore and molecules into a lattice.

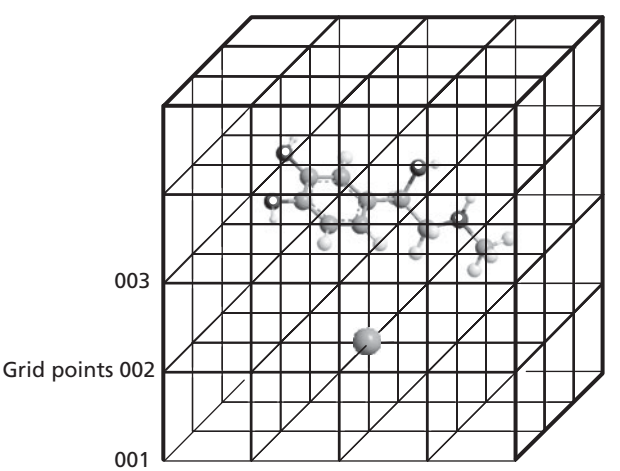
18.10.2 Relating shape and electronic distribution to biological activity

Defining the steric and electrostatic fields of a series of molecules is relatively straightforward and is carried out automatically by the software program. The next stage is to relate these properties to the biological activity of the molecules. This is less straightforward and differs significantly from traditional QSAR. In traditional QSAR, there are relatively few variables involved. For example, if we consider log P, π, σ, and a size factor for each molecule, then we have four variables per molecule to compare against biological activity. With 100 molecules in the study, there are far more molecules than variables and it is possible to come up with an equation relating variables to biological activity, as described previously.

In 3D QSAR, the variables for each molecule are the calculated steric and electronic interactions at a couple of thousand lattice points. With 100 molecules in the study, the number of variables now far outweighs the number of structures, and it is not possible to relate these to biological potency by the standard multiple linear regression analysis described in section 18.1. A different statistical procedure has to be followed using a technique called **partial least squares** (PLS). Essentially, it is an analytical computing process which is repeated over and over again (iterated) to try to find the best formula relating biological property against the various variables. As part of the process, the number of variables is reduced as the software filters out those which are clearly unrelated to biological activity.

An important feature of the analysis is that a structure is deliberately left out as the computer strives to form some form of relationship. Once a formula has been defined, the formula is tested against the structure which was left out. This is called **cross-validation** and tests how well the formula predicts the biological property for the molecule which was left out. The results of this are fed back into another round of calculations, but now the structure which was left out is included in the calculations and a different structure is left out. This leads to a new improved formula which is once again tested against the compound that was left out, and so the process continues until cross validation has been carried out against all the structures.

At the end of the process, the final formula is obtained (Fig. 18.22). The predictability of this final



Probe atom

Position probe atom at each grid point
Measure steric and electrostatic fields

Tabulate fields for each compound at each grid point

Compound	Biological activity	Steric fields (S) at grid points (001-998)					Electrostatic fields (E) at grid points (001-098)				
		S001	S002	S003	S004	S005 etc	E001	E002	E003	E004	E005 etc
1	5.1										
2	6.8										
3	5.3										
4	6.4										
5	6.1										

Partial least squares analysis (PLS)

QSAR equation Activity = aS001 + bS002 +……..mS998 + nE001 +…….+yE998 + z

FIGURE 18.22 Measuring steric and electronic fields.

equation is quantified by the **cross-validated correlation coefficient** r^2, which is usually referred to as q^2. In contrast to normal QSAR, where r^2 should be greater than 0.8, values of q^2 greater than 0.3 are considered significant. It is more useful, though, to give a graphical representation showing which regions around the molecule are important to biological activity on steric or electronic grounds. Therefore, a steric map shows a series of coloured contours indicating beneficial and detrimental steric interactions around a representative molecule from the set of molecules tested (Fig. 18.23). A similar contour map is created to illustrate electrostatic interactions.

An example of a 3D QSAR study is described in the case study described in section 18.10.6.

For additional material see Web article 5: The design of a serotonin antagonist as a possible anxiolytic agent

18.10.3 Advantages of CoMFA over traditional QSAR

Some of the problems involved with a traditional QSAR study include the following:

- only molecules of similar structure can be studied;
- the validity of the numerical descriptors is open to doubt. These descriptors are obtained by measuring reaction rates and equilibrium constants in model reactions and are listed in tables. However, separating one property from another is not always possible in experimental measurement. For example, the Taft steric factor is not purely a measure of the steric factor, because the measured reaction rates used to define it are also affected by electronic factors. Also, the *n*-octanol/water partition coefficients which are used to measure log *P* are known to be affected by the hydrogen bonding character of molecules;
- the tabulated descriptors may not include entries for unusual substituents;
- it is necessary to synthesize a range of molecules where substituents are varied in order to test a particular property (e.g. hydrophobicity). However, synthesizing such a range of compounds may not be straightforward or feasible;
- traditional QSAR equations do not directly suggest new compounds to synthesize.

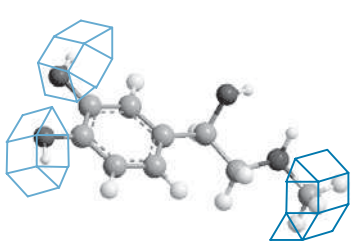


FIGURE 18.23 Definition of favourable and unfavourable interactions around a representative molecule.

These problems are avoided with CoMFA which has the following advantages:

- favourable and unfavourable interactions are represented graphically by 3D contours around a representative molecule. A graphical picture such as this is easier to visualize than a mathematical formula;
- in CoMFA, the properties of the test molecules are calculated individually by computer programs. There is no reliance on experimental or tabulated factors. There is no need to confine the study to molecules of similar structure. As long as one is confident that all the compounds in the study share the same pharmacophore and interact in the same way with the target, they can all be analysed in a CoMFA study;
- the graphical representation of beneficial and non-beneficial interactions allows medicinal chemists to design new structures. For example, if a contour map shows a favourable steric effect at one particular location, this implies that the target binding site has space for further extension at that location. This may lead to further favourable receptor–drug interactions;
- both traditional and 3D QSAR can be used without needing to know the structure of the biological target.

18.10.4 Potential problems of CoMFA

There are several potential problems in using CoMFA:

- it is important to know the active conformation for each of the molecules in the study. Identifying the active conformation is easy for rigid structures such as steroids, but it is more difficult for flexible molecules that are capable of several bond rotations. Therefore, it is useful to have a conformationally restrained analogue which is biologically active and which can act as a guide to the likely active conformation. More flexible molecules can then be constructed on the computer with the conformation most closely matching that of the more rigid analogue. If the structure of the target binding site is known, it can be useful in deciding the likely active conformations of molecules;
- each molecule in the study must be positioned correctly in the grid so that it is properly aligned with respect to all the other molecules. A common pharmacophore can be used to aid this process as described earlier. However, it may be difficult to identify the

pharmacophore in some molecules. In that case, a pharmacophore mapping exercise could be carried out (section 17.11). This is likely to be successful if there are some rigid active compounds among the compounds being studied. An alternative method of alignment is to align the molecules based on their structural similarity. This can be done automatically using what is known as 'topomer' methodology;

- one has to be careful to ensure that all the compounds in the study interact with the target in similar ways. For example, a 3D QSAR study on all possible **acetylcholinesterase** inhibitors is doomed to failure. Firstly, the great diversity of structures involved makes it impossible to align these structures in an unbiased way or to generate a 3D pharmacophore. Secondly, the various inhibitors do not interact with the target enzyme in the same way. X-ray crystallographic studies of enzyme–inhibitor complexes show that the inhibitors **tacrine**, **edrophonium**, and **decamethonium** all have different binding orientations in the active site;
- 3D QSAR provides a summary of how structural changes in a drug affect biological activity, but it is dangerous to assume too much. For example, a 3D QSAR model may show that increasing the bulk of the molecule at a particular location increases activity. This might suggest that there is an accessible hydrophobic pocket allowing extra binding interactions. However, it is possible that the extra steric bulk causes the molecule to bind in a different orientation from the other molecules in the analysis and that this is the reason for the increased activity;
- it has been found that slightly different orientations of the grid from one study to another can produce different results for the same set of compounds.

18.10.5 Other 3D QSAR methods

CoMFA continues to be the most popular program for studies into 3D QSAR, but it does suffer a number of disadvantages, as described above. Two of the more serious problems are the spurious results that can be obtained if compounds are not properly aligned, or if the orientation of the grid box is slightly different between studies. Users of the program can also choose different spacings between the grid points, and this can give poor results if the grid is too coarse or too fine. The method is also computationally expensive, requiring a lot of calculations for each molecule in the study, and so powerful computers are needed to cope with the huge memory requirement.

Other 3D QSAR programs have been developed in an attempt to address some of these issues. Examples include **HINT**, which can be used alongside CoMFA to measure a hydrophobic field, **CoMSIA** which includes hydrogen bonding and hydrophobic fields, as well as steric and electrostatic fields, and **CoMASA**, which uses fewer calculations.

Some 3D QSAR programs use the intrinsic molecular properties of compounds rather than using a probe to measure the property fields surrounding them. Four examples are **SOMFA**, **HASL**, **CoMMA**, and **MS-Whim**. Other programs are used to model hypothetical pseudo-receptors. These include **Quasar**, **WeP**, and **GRIND**.

KEY POINTS

- CoMFA is an example of a 3D QSAR program which measures steric and electrostatic fields round a series of structures and relates these to biological activity.
- A comparison of the steric and electrostatic fields for different molecules against their biological activity allows the definition of steric and electrostatic interactions which are favourable and unfavourable for activity. These can be displayed visually as contour lines.
- It is necessary to define the active conformation and pharmacophore for each molecule in a CoMFA study. Alignment of the molecules is crucial.
- Unlike conventional QSAR studies, molecules of different structural classes can be compared if they share the same pharmacophore.
- 3D QSAR does not depend on experimentally measured parameters.
- A variety of different 3D QSAR programs have been developed.

18.10.6 Case study: inhibitors of tubulin polymerization

Colchicine (Fig. 18.24) is a lead compound for agents which act as inhibitors of tubulin polymerization (section 2.7.1) that might be useful in the treatment of arthritis. Other lead compounds have been discovered which bind to tubulin at the same binding site, and so a study was carried out to compare the various structural classes interacting in this way. In this 3D QSAR study, 104 such

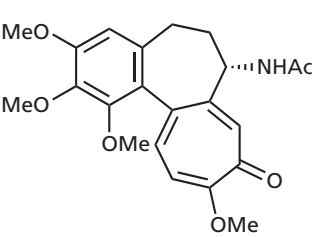


FIGURE 18.24 Colchicine.

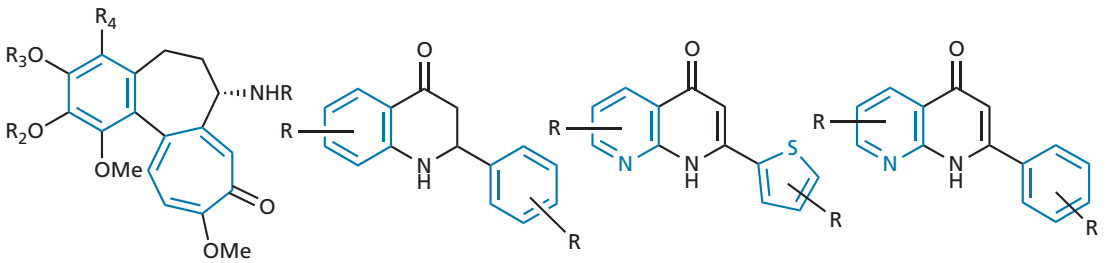


FIGURE 18.25 Structural classes used in the 3D QSAR study.

agents were tested, belonging to four distinct families of compounds (Fig. 18.25); 51 compounds were used as a 'training set' for the analysis itself and 53 were used as a 'testing set' to test the predictive value of the results. Both sets contained a mixture of structural classes having both low and high activity.

The first task was to work out how to align these different classes of molecule. Colchicine is the most rigid of the four and also has a high affinity for tubulin. Therefore, it was chosen as the template on to which the other structures would be aligned. The relevant pharmacophore in colchicine was identified as the two aromatic rings. Molecular modelling was now carried out on each of the remaining structures to generate various conformations. Each conformation was compared with colchicine to find the one that would allow the pharmacophores in each structure to be aligned. This was then identified as the active conformation.

Once the active conformations for each structure had been identified, they were fitted into the lattice of grids described previously such that each structure was properly aligned. The steric and electrostatic fields round each molecule were calculated using a probe atom, then the 3D QSAR analysis was carried out to relate the fields to the measured biological activity.

The results of the 3D QSAR analysis are summarized as contour lines around a representative molecule (Fig. 18.26). For the steric interactions, solid contours represent fields that are favourable for activity and the dashed lines show fields that are unfavourable. For the electrostatic interactions, solid lines are regions where positively charged species improve affinity, and dashed lines indicate regions where negatively charged groups are favourable.

The results revealed that introducing steric bulk around the aromatic ring is more crucial to activity than introducing steric bulk around the bicyclic system. Based on this evidence, the structure shown in Fig. 18.27 was synthesized. The predicted value of pIC_{50} for the compound was 5.62. The actual value was in close agreement at 6.04. ($pIC_{50} = -\log [IC_{50}]$ where IC_{50} is the concentration of inhibitor required to produce 50% enzyme inhibition).

The steric fields of the Tripos' CoMFA analysis (Fig. 18.26) were subsequently placed into a model of the binding site. It was found that the bad steric regions were in the same regions as the peptide backbone, whereas the favourable steric areas were in empty spaces.

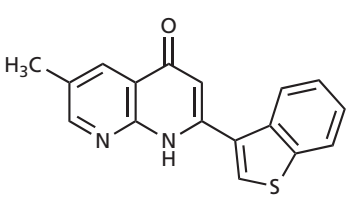


FIGURE 18.27 Novel agent designed on the basis of the 3D QSAR study.

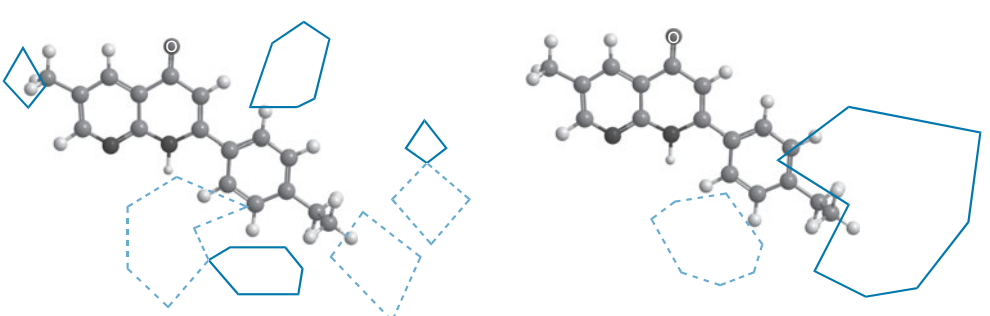


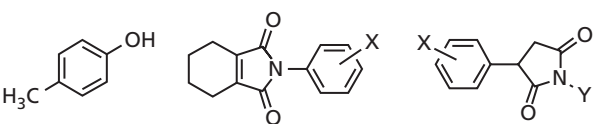
FIGURE 18.26 Results of the 3D QSAR analysis ($q^2 = 0.637$).

QUESTIONS

1. Using values from Table 18.1, calculate the log *P* value for structure (I) (log *P* for benzene = 2.13).
2. Several analogues of a drug are to be prepared for a QSAR study which will consider the effect of various aromatic substituents on biological activity. You are asked whether the substituents (SO_2NH_2, CF_3, CN, CH_3SO_2, SF_3, $CONH_2$, OCF_3, CO_2H, Br, I) are relevant to the study. What are your thoughts?
3. A lead compound has a monosubstituted aromatic ring present as part of its structure. An analogue was synthesized containing a *para*-chloro substituent which had approximately the same activity. It was decided to synthesize an analogue bearing a methyl group at the *para*-position. This showed increased activity. What analogue would you prepare next and why?
4. The following QSAR equation was derived for the pesticide activity of structure (II). Explain what the various terms mean and whether the equation is a valid one. Identify what kind of substituents would be best for activity.

   $\log 1/C = 1.08\pi_x + 2.41F_x + 1.40\ R_x - 0.072\ MR_x + 5.25$ ($n = 16$, $r^2 = 0.840$, $s = 0.59$)
5. A QSAR equation for the anticonvulsant (III) was derived as follows:

   $\log 1/C = 0.92\ \pi_x - 0.34\ \pi_x^2 + 3.18$ ($n = 15$, $r^2 = 0.902$, $s = 0.09$, $\pi_o = 1.35$). What conclusions can you draw from this equation? Would you expect activity to be greater if X = CF_3 rather than H or CH_3?
6. The following QSAR equation is related to the mutagenic activity of a series of nitrosoamines; $\log 1/C = 0.92\ \pi + 2.08\ \sigma - 3.26$ ($n = 12$, $r^2 = 0.794$, $s = 0.314$). What sort of substituent is likely to result in high mutagenic activity?



I II III

FURTHER READING

Craig, P. N. (1971) Interdependence between physical parameters and selection of substituent groups for correlation studies. *Journal of Medicinal Chemistry* **14**, 680–684.

Cramer, R. D. (2003) Topomer CoMFA: A design methodology for rapid lead optimization. *Journal of Medicinal Chemistry* **46**, 374–388.

Cramer, R. D. III, Snader, K. M., Willis, C. R., Chakrin, L. W., Thomas, J., and Sutton, B. M. (1979) Application of quantitative structure–activity relationships in the development of the antiallergenic pyranenamines. *Journal of Medicinal Chemistry* **22**, 714–725.

Cramer, R. D., Patterson, D. E., and Bunce, J. D. (1988) Comparative field analysis (CoMFA). *Journal of the American Chemical Society* **110**, 5959–5967.

Hansch, C. and Leo, A. (1995) *Exploring QSAR*. American Chemical Society, Washington, DC.

Kellogg, G. E. and Abraham, D. J. (2000) Hydrophobicity: Is $LogP_{o/w}$ more than the sum of its parts? *European Journal of Medicinal Chemistry* **35**, 651–661.

Kotani, T. and Higashiura, K. (2004) Comparative molecular active site analysis (CoMASA). 1. An approach to rapid evaluation of 3D QSAR. *Journal of Medicinal Chemistry* **47**, 2732–2742.

Kubini, H., Folkers, G., and Martin, Y. C. (eds) (1998) *3D QSAR in Drug Design*. Kluwer/Escom, Dordrecht.

Martin, Y. C. and Dunn, W. J. (1973) Examination of the utility of the Topliss schemes by analog synthesis. *Journal of Medicinal Chemistry* **16**, 578–579.

Sutherland, J. J., O'Brien, L. A., and Weaver, D. F. (2004) A comparison of methods for modeling quantitative structure-activity relationships. *Journal of Medicinal Chemistry* **47**, 5541–5554.

Taft, R. W. (1956) Separation of polar, steric and resonance effects in reactivity. In: Newman, M. S. (ed.) *Steric Effects in Organic Chemistry*. John Wiley and Sons, New York.

Topliss, J. G. (1972) Utilization of operational schemes for analog synthesis in drug design. *Journal of Medicinal Chemistry* **15**, 1006–1011.

Verloop, A., Hoogenstraaten, W., and Tipker, J. (1976) Development and application of new steric substituent parameters in drug design. *Medicinal Chemistry* **11**, 165–207.

van de Waterbeemd, H., Testa, B., and Folkers, G. (eds) (1997) *Computer-assisted Lead Finding and Optimization*. Wiley-VCH, New York.

Zhang, S. X., Feng, J., Kuo, S. C., Brossi, A., Hamel, E. Tropsha, A., and Lee, K. H. (2000) Antitumor agents. 199. Three–dimensional quantitative structure–activity relationship study of the colchicine binding site ligands using comparative molecular field analysis. *Journal of Medicinal Chemistry* **43**, 167–176.

Titles for general further reading are listed on p. 763.

CASE STUDY 5

Design of a thymidylate synthase inhibitor

In this case study, we shall look at an early example of how the strategies of *de novo* drug design (section 17.15) and structure-based drug design (section 13.3.11) were used together to develop an active compound that went forward for clinical trials. The research in question involved the design of inhibitors for the enzyme **thymidylate synthase** (section 21.3.2). This enzyme catalyses the methylation of **deoxyuridylate monophosphate** (**dUMP**) to **deoxythymidylate monophosphate** (**dTMP**) using **5,10-methylenetetrahydrofolate** as a coenzyme (Fig. CS5.1). Inhibitors of this enzyme have been shown to be anti-tumour agents which prevent the biosynthesis of one of the required building blocks for DNA. Traditional inhibitors have been modelled on dUMP or the enzyme cofactor 5,10-methylenetetrahydrofolate (Fig. CS5.2), which means that these inhibitors are structurally related to the natural substrate and cofactor. Unfortunately, this increases the possibility of side effects resulting from inhibition of other enzymes and receptors which use these molecules as natural ligands. Therefore, it was decided that *de novo* drug design would be used to design a novel structure which was unrelated to either of the natural substrates.

Before starting the *de novo* design a good supply of the enzyme was required. Although human thymidylate synthase was not readily available in large quantities, it was possible to obtain good quantities of the bacterial version from *Escherichia coli* by using recombinant DNA technology to clone the gene and then expressing it in fast-growing cells (section 6.4). The bacterial enzyme is not identical to the human version, but it is very similar and so it was considered a reasonable analogue. The enzyme was crystallized along with the known inhibitors **5-fluorodeoxyuridylate** and **CB 3717** (Fig. CS5.3). These structures mimic the substrate and the coenzyme, respectively, and bind to the sites normally occupied by these structures. The structure of the enzyme–inhibitor complex was then determined by X-ray crystallography and downloaded on to a computer.

A study of the enzyme–inhibitor complex revealed where the inhibitors were bound and also the binding interactions involved. For CB 3717, the binding interactions around the pteridine portion of the inhibitor were identified as involving hydrogen bonding interactions to two amino acids (the carboxylate ion of Asp-169 and the main chain peptide link next to Ala-263). There was also a hydrogen bonding interaction to a water molecule which acted as a hydrogen bonding bridge to Arg-21 (Fig. CS5.4). Using molecular modelling, the inhibitor was deleted from the binding site to allow further analysis of the empty binding site. Generating the empty binding site from the enzyme–ligand complex is better than studying the empty binding site from the pure enzyme because the latter does not take into account the induced fit that occurs on ligand binding.

A grid was set up within the binding site and an aromatic CH probe was placed at each grid point to measure hydrophobic interactions and thus identify hydrophobic regions (see section 17.7.5). From this analysis, it was

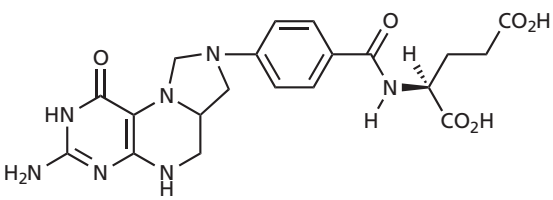


FIGURE CS5.2 5,10-Methylenetetrahydrofolate.

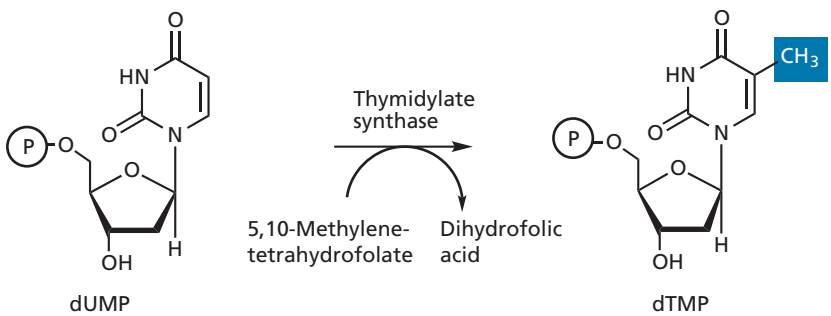


FIGURE CS5.1 Reaction catalysed by thymidylate synthase. (Ⓟ = Phosphate)

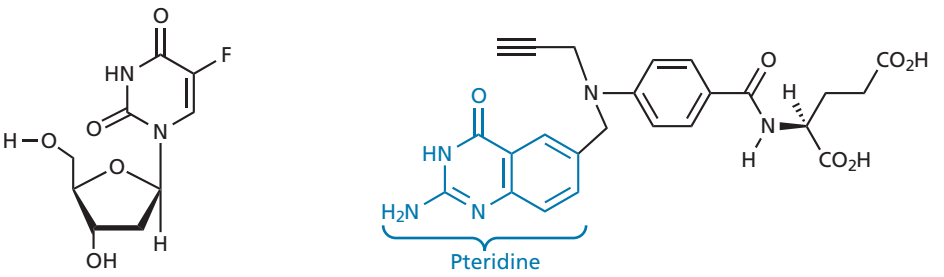


FIGURE CS5.3 Inhibitors of thymidylate synthase.

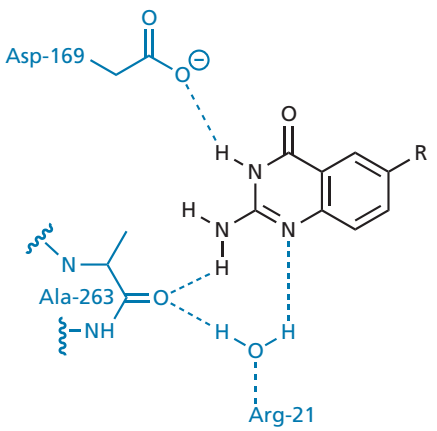


FIGURE CS5.4 Binding interactions holding a pteridine moiety in the active site.

discovered that the pteridine portion of CB 3717 was positioned in a hydrophobic pocket despite the presence of the hydrogen bonding interactions which held it there. The boundaries of this hydrophobic region were determined and a naphthalene ring was found to be a suitable hydrophobic molecule to fit the pocket, yet still leave room for the addition of a functional group which would be capable of forming the important hydrogen bonds.

The functional group chosen was a cyclic amide which was fused to the naphthalene scaffold to create a naphthostyryl scaffold (Fig. CS5.5). Modelling suggested that the NH portion of the amide would bind to Asp-169 while the carbonyl group would bind to the water molecule identified above. A substituent was now added to the naphthostyryl scaffold in order to gain access to the space normally occupied by the benzene ring of the cofactor. A dialkylated amine was chosen as the linking unit and was placed at position 5 of the structure. There were several reasons for this. Firstly, adding an amine at this position was easy to carry out synthetically. Secondly, the two substituents on the amine could be easily varied, which would allow fine-tuning of the compound. Lastly, by using an amine it would be possible to have a branching point which could have different substituents without adding an asymmetric centre. If a carbon atom had been added instead, two different substituents would have led to an asymmetric centre with all the attendant complications that would entail (section 13.3.8).

Modelling demonstrated that a benzyl group was a suitable substituent for the amine in order to access the space normally occupied by the benzene ring of the cofactor. A phenylsulphonyl piperazine group was then added to the *para* position of the aromatic ring in order

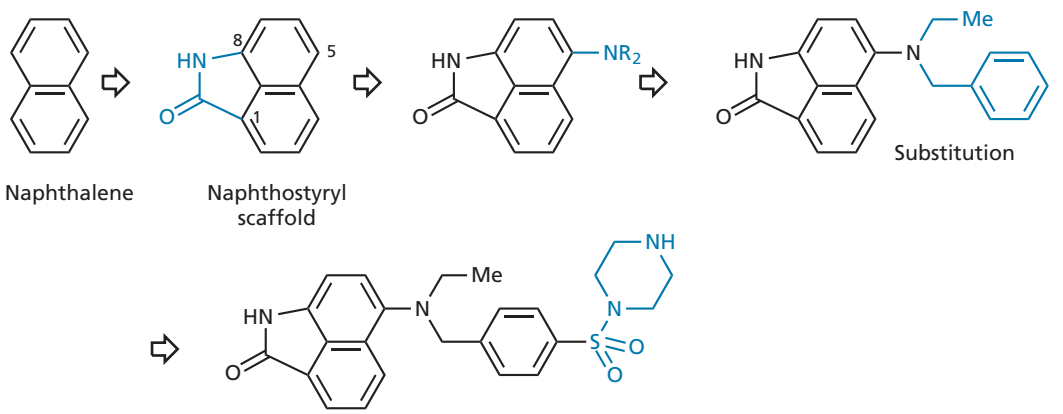


FIGURE CS5.5 Design of an enzyme inhibitor by *de novo* methods.

to make the molecule more water soluble—a necessary property if the synthesized structures were to be bound to the enzyme and crystallized for further X-ray crystallographic studies. The positioning of this group was important because the modelling studies showed that it would protrude from the binding site and still make contact with surrounding water. This meant that the group would not have to be desolvated for the drug to bind—a process which would involve an energy penalty (section 1.3.6).

This structure was now synthesized and was found to inhibit both the bacterial and human versions of the enzyme, with higher activity for the human enzyme. This represents the successful *de novo* design of a novel lead structure.

It was now time to move to structure–based drug design such that the binding interactions and activity of the lead compound could be optimized.

A crystal structure of the novel inhibitor bound to the bacterial enzyme was successfully obtained and studied to see whether the inhibitor had fitted the binding site as expected. In fact, it was found that the naphthalene ring of the inhibitor was wedged deeper into the pocket than expected because of more favourable hydrophobic interactions. As a result, the cyclic amide failed to form the direct hydrogen bond interaction to Asp-169 which had been planned and was hydrogen bonding to a bridging water molecule instead. The lactam carbonyl oxygen was also too close to Ala-263, which had caused this residue to shift 1 Å from its usual position. This, in turn, had displaced the water molecule, which had been the intended target for hydrogen bonding (Fig. CS5.6).

By studying the actual position of the structure in the binding site, it was possible to identify four areas where extra substituents could fill up empty space and perhaps improve binding. These are shown in Fig. CS5.7.

Various structures were proposed then overlaid on the lead compound (still docked within the binding site) to see whether they fitted the binding site. Only those which passed this test and were in stable conformations were synthesized and tested for activity (41 in total). The optimum substituent at each position was then identified.

- In region 1 (R^1), modelling showed that this substituent fitted into a hydrophobic pocket that became hydrophilic the deeper one got. This suggested that a hydrogen bonding substituent at the end of an alkyl chain might be worth trying (the *extension strategy*; section 13.3.2) and, indeed, a CH_2CH_2OH group led to an improvement in binding affinity. It was also found that a methyl group was better than the original ethyl group.
- In region 2 (R^2), the carbonyl oxygen was replaced by an amidine group which would be capable of hydrogen bonding to the carbonyl oxygen of Ala-263, rather than repelling it. An added advantage in using a basic amidine group was the fact that there was a good chance that it would become protonated, allowing a stronger ionic interaction with Asp-169, as well as a better hydrogen bonding interaction with Ala-263. When this structure was synthesized, it was found to

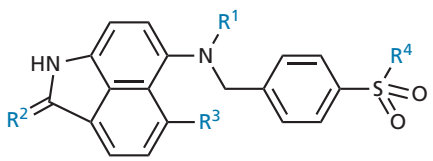


FIGURE CS5.7 Variable positions (in colour).

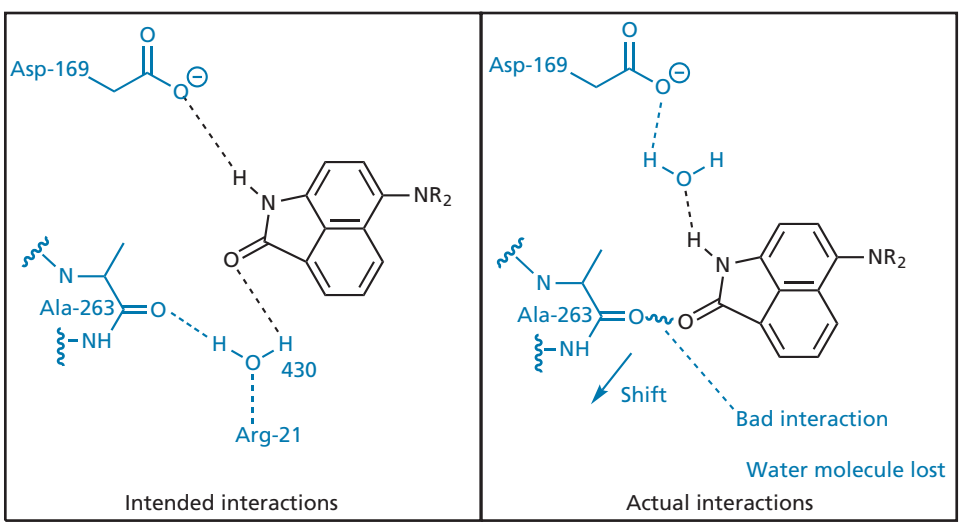


FIGURE CS5.6 Intended versus actual interactions of the inhibitor with the active site.

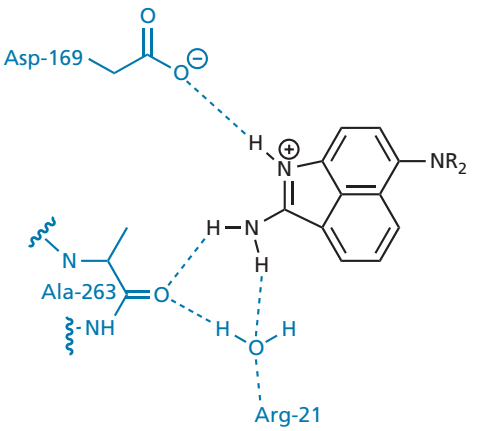


FIGURE CS5.8 Binding interactions of the modified inhibitor with the active site.

have improved inhibition, and a crystal structure of the enzyme–inhibitor complex showed that the expected interactions were taking place (Fig. CS5.8). Moreover, Ala-263 had returned to its original position, permitting the return of the bridging water molecule.

- In region 3 (R^3), there was room for a small group, such as a chlorine atom or a methyl group; both of these substituents led to an increase in activity.
- Region 4 (R^4) was relatively unimportant for inhibitory activity, as groups at this position protrude out of the active site into the surrounding solvent and have only minimal contact with the enzyme. Nevertheless, the piperazine ring was replaced by a morpholine group because the latter had some advantages with respect to selectivity and pharmacological properties.

Having identified the optimum groups at each position, structures were synthesized combining some, or all, of these groups. The presence of the amidine resulted in the best improvement in activity and so the presence of this group was mandatory. Interestingly, adding all the optimum groups is not as beneficial as adding some of them.

The modified structure (Fig. CS5.9) was synthesized and was found to be a potent inhibitor, which was 500 times more active than the original amide. A crystal structure of the enzyme–inhibitor complex showed a much better fit and the compound was put forward for clinical trials as an anti-tumour agent.

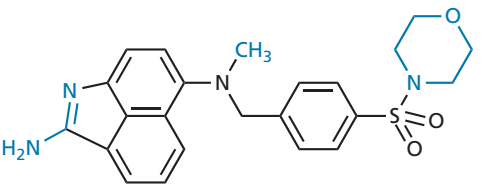


FIGURE CS5.9 Modified inhibitor put forward for clinical trials.

The case study illustrates many of the general principles behind *de novo* design that were described in section 17.15.1. For example, the designed lead compound was fairly flexible and did not fill up all the space available in the binding site. As we have seen, a different binding mode took place from that predicted, and this might not have occurred if a more rigid and more closely fitting structure had been designed. The fact that binding did take place allowed structure-based drug design to be carried out on the information obtained.

The importance of synthetic feasibility was considered throughout the process and was one of the reasons for introducing an amine substituent to the naphthostyryl ring system.

Finally, it is important to appreciate that binding studies using target proteins from different species may produce slightly different results. For example, the computer modelling studies described above were carried out on bacterial thymidylate synthase enzyme rather than the human version. Fortunately, the activities of the designed inhibitors were actually greater for the human enzyme than for the bacterial enzyme, and this was put down to the fact that the hydrophobic space available for the naphthalene ring was larger in the human enzyme than in the bacterial one. Fortunately, most changes carried out had beneficial effects for both enzymes, with one exception; adding a methyl group at R^3 of the amidine led to an increase in activity for the bacterial enzyme, but not for the human enzyme.

FURTHER READING

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 drug design. *Journal of Medicinal Chemistry* **37**, 1035–1054.

19 Antibacterial agents

The fight against bacterial infections over the last 70 years has been one of the great success stories of medicinal chemistry, yet it remains to be seen whether it will last. Bacteria, such as *Staphylococcus aureus*, have the worrying ability to gain resistance to known drugs and so the search for new drugs is never-ending. Although deaths from bacterial infection have dropped in the developed world, bacterial infection is still a major cause of death in the developing world. For example, the World Health Organization estimated that tuberculosis was responsible for about 2 million deaths in 2002 and that 1 in 3 of the world's population was infected. The same organization estimated that in the year 2000, 1.9 million children died worldwide of respiratory infections with 70% of these deaths occurring in Africa and Asia. They also estimated that, each year, 1.4 million children died from gut infections and the diarrhoea resulting from these infections. In the developed world, deaths from food poisoning due to virulent strains of *Escherichia coli* have attracted widespread publicity, while tuberculosis has returned as a result of the AIDS epidemic.

The topic of antibacterial agents is a large one and terms are used in this chapter which are unique to this particular field. Rather than clutter the text with explanations and definitions, Appendix 5 contains explanations of such terms as aerobic and anaerobic organisms; antibacterial and antibiotic substances; **cocci**; **bacilli**; streptococci; and staphylococci. Appendix 5 also explains briefly the difference between bacteria, algae, protozoa, and fungi. The emphasis in this chapter is on agents that act against bacteria, but some of those described also act against protozoal infections and this may be mentioned in the text.

19.1 History of antibacterial agents

There is evidence of antibacterial herbs or potions being used for many centuries. For example, the Chinese used mouldy soybean curd to treat carbuncles, boils, and other infections. Greek physicians used wine, myrrh, and inorganic salts. In the Middle Ages, certain types of honey were used to prevent infections following arrow wounds. Of course in those days, there was no way of knowing that bacteria were the cause of these infections.

Bacteria are single-cell microorganisms first identified in the 1670s by van Leeuwenhoek, following his invention of the microscope. It was not until the nineteenth century, however, that their link with disease was appreciated. This followed the elegant experiments carried out by the French scientist Pasteur, who demonstrated that specific bacterial strains were crucial to fermentation and that these, and other, microorganisms were more widespread than was previously thought. The possibility that these microorganisms might be responsible for disease began to take hold.

An early advocate of a 'germ theory of disease' was the Edinburgh surgeon Lister. Despite the protests of several colleagues who took offence at the suggestion that they might be infecting their own patients, Lister introduced **carbolic acid** as an antiseptic and sterilizing agent for operating theatres and wards. The improvement in surgical survival rates was significant.

During the latter half of the nineteenth century, scientists such as Koch were able to identify the microorganisms responsible for diseases such as tuberculosis, cholera, and typhoid. Methods of **vaccination** were studied and research was carried out to try and find effective antibacterial agents or antibiotics. The scientist who can lay claim to be the father of chemotherapy—the use of chemicals against infection—was Paul Ehrlich. Ehrlich spent much of his career studying histology, then immunochemistry, and won a Nobel prize for his contributions to immunology. In 1904, however, he switched direction and entered a field which he defined as chemotherapy. Ehrlich's **principle of chemotherapy** was that a chemical could directly interfere with the proliferation of microorganisms at concentrations tolerated by the host. This concept was popularly known as the **magic bullet**, where the chemical was seen as a bullet which could

search out and destroy the invading microorganism without adversely affecting the host. The process is one of **selective toxicity**, where the chemical shows greater toxicity to the target microorganism than to the host cells. Such selectivity can be represented by a **chemotherapeutic index**, which compares the minimum effective dose of a drug with the maximum dose that can be tolerated by the host. This measure of selectivity was eventually replaced by the currently used **therapeutic index**.

By 1910, Ehrlich had successfully developed the first example of a purely synthetic antimicrobial drug. This was the arsenic-containing compound **salvarsan** (Fig. 19.1). Although it was not effective against a wide range of bacterial infections, it did prove effective against the protozoal disease of sleeping sickness (trypanosomiasis) and the spirochete disease of syphilis. The drug was used until 1945 when it was replaced by penicillin (see also Box 19.20).

Over the next 20 years, progress was made against a variety of protozoal diseases, but little progress was made in finding antibacterial agents until the introduction in 1934 of **proflavine** (Fig. 19.1)—a drug which was used during World War II against bacterial infections in deep surface wounds. Unfortunately, it was too toxic to be used against systemic bacterial infections (i.e. those carried in the bloodstream) and there was still an urgent need for agents which would fight these infections.

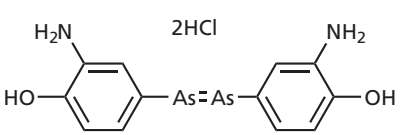
This need was answered in 1935 when it was discovered that a red dye called **prontosil** was effective against streptococcal infections *in vivo*. As discussed later, prontosil was recognized eventually as a prodrug for a new class of antibacterial agents—the **sulpha drugs** or **sulphonamides**. The discovery of these drugs was a real breakthrough, as they represented the first drugs to be effective against systemic bacterial infections. In fact, they were the only effective drugs until penicillin became available in the early 1940s.

Although **penicillin** was discovered in 1928, it was not until 1940 that effective means of isolating it were developed by Florey and Chain. Society was then rewarded with a drug which revolutionized the fight against bacterial infection and proved even more effective than the sulphonamides. Despite penicillin's success, it was not effective against all types of infection and the need for new antibacterial agents still remained. Penicillin is an example of a toxic fungal metabolite that kills bacteria and allows the fungus to compete for nutrients. The realization that fungi might be a source for novel antibiotics spurred scientists into a huge investigation of microbial cultures from all round the globe.

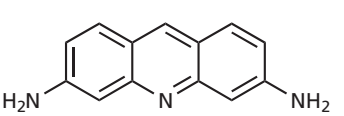
In 1944, the antibiotic **streptomycin** was discovered from a systematic search of soil organisms. It extended the range of chemotherapy to the tubercle bacillus and a variety of Gram-negative bacteria. This compound was the first example of a series of antibiotics known as the **aminoglycoside** antibiotics. After World War II, the search continued leading to the discovery of **chloramphenicol** (1947), the peptide antibiotics (e.g. **bacitracin**, 1945), the **tetracycline** antibiotics (e.g. **chlortetracycline**, 1948), the macrolide antibiotics (e.g. **erythromycin**, 1952), the cyclic peptide antibiotics (e.g. **valinomycin**), and the first example of a second major group of β-lactam antibiotics, **cephalosporin C** (1955).

As far as synthetic agents were concerned, **isoniazid** was found to be effective against human tuberculosis in 1952, and in 1962 **nalidixic acid** (the first of the **quinolone** antibacterial agents) was discovered. A second-generation of this class of drugs was introduced in 1987 with **ciprofloxacin**.

Many antibacterial agents are now available and the vast majority of bacterial diseases have been brought under control (e.g. syphilis, tuberculosis, typhoid, bubonic plague, leprosy, diphtheria, gas gangrene, tetanus, and gonorrhoea). This represents a great achievement for medicinal chemistry and it is perhaps sobering to consider the hazards society faced in the days before penicillin. Septicaemia was a risk faced by mothers during childbirth and could lead to death. Ear infections were common, especially in children, and could lead to deafness. Pneumonia was a frequent cause of death in hospital wards. Tuberculosis was a major problem, requiring special isolation hospitals built away from populated centres. A simple cut or a wound could lead to severe infection requiring the amputation of a limb, while the threat of peritonitis lowered the success rates of surgical operations. This was in the 1930s—still within living memory for many. Perhaps those of us born since World War II take the success of antibacterial agents too much for granted.



Salvarsan



Proflavine

FIGURE 19.1 Salvarsan and proflavine. (The structure of salvarsan shown here is a simplification; it is, in fact, a cyclic trimer with no As = As bonds.)

19.2 The bacterial cell

The success of antibacterial agents owes much to the fact that they can act selectively against bacterial cells rather than animal cells. This is largely because bacterial and animal cells differ both in their structure and in their biosynthetic pathways. Let us consider some of the differences between the bacterial cell (defined as **prokaryotic**) (Fig. 19.2) and the animal cell (defined as **eukaryotic**).

Differences between bacterial and animal cells:

- the bacterial cell does not have a defined nucleus, whereas the animal cell does;
- animal cells contain a variety of structures called organelles (mitochondria, endoplasmic reticulum, etc.), whereas the bacterial cell is relatively simple;
- the biochemistry of a bacterial cell differs significantly from that of an animal cell. For example, bacteria may have to synthesize essential vitamins which animal cells can acquire intact from food. The bacterial cells must have the enzymes to catalyse these reactions. Animal cells do not, because the reactions are not required;
- the bacterial cell has a cell membrane and a cell wall, whereas the animal cell has only a cell membrane. The cell wall is crucial to the bacterial cell's survival. Bacteria have to survive a wide range of environments and osmotic pressures, whereas animal cells do not. If a bacterial cell lacking a cell wall was placed in an aqueous environment containing a low concentration of salts, water would freely enter the cell as a result of osmotic pressure. This would cause the cell to swell and eventually burst. The scientific term for this is **lysis**. The cell wall does not stop water flowing into the cell directly, but it does prevent the cell from swelling and so indirectly prevents water entering the cell. Bacteria can be characterized by a staining technique which allows them to be defined as **Gram-positive** or **Gram-negative** (Appendix 5). Bacteria with a thick cell wall (20–40 nm) are stained purple and defined as Gram-positive. Bacteria with a thin cell wall (2–7 nm) are stained pink and are defined as Gram-negative. Although Gram-negative bacteria have a thin cell wall, they have an additional outer membrane not present in Gram-positive bacteria. This outer membrane is made up of lipopolysaccharides—similar in character to the cell membrane. These differences in cell walls and membranes have important consequences for the different vulnerabilities of Gram-positive and Gram-negative bacteria to antibacterial drugs.

19.3 Mechanisms of antibacterial action

There are five main mechanisms by which antibacterial agents act (Fig. 19.2).

- *Inhibition of cell metabolism*: antibacterial agents which inhibit cell metabolism are called **antimetabolites**. These compounds inhibit the metabolism of a microorganism, but not the metabolism of the host. They can do this by inhibiting an enzyme-catalysed reaction which is present in the bacterial cell, but not in animal cells. The best-known examples of antibacterial agents acting in this way are the sulphonamides. It is also possible for antibacterial agents to show selectivity against enzymes which are present in both the bacterial and mammalian cell, as long as there are significant differences in structure between the two.

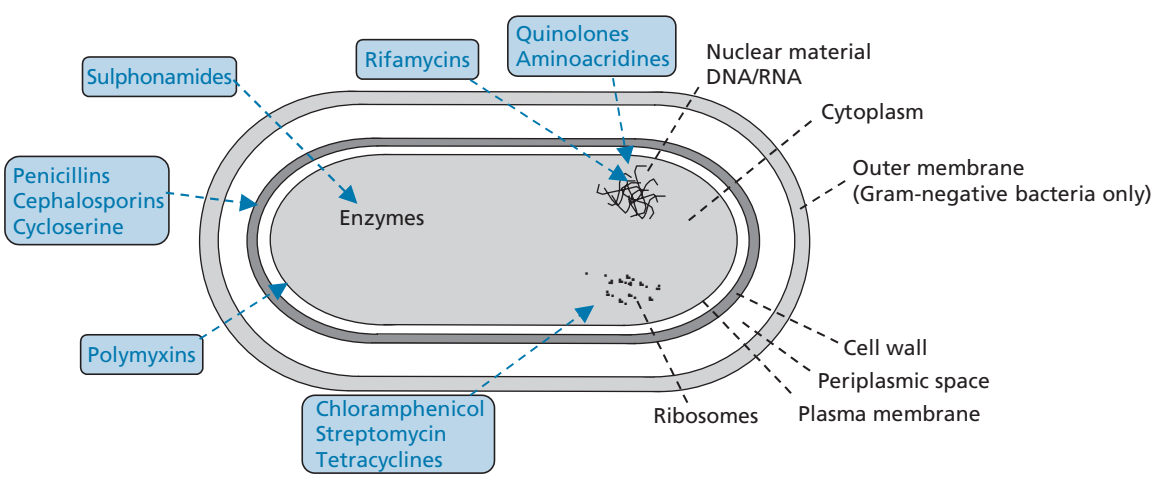


FIGURE 19.2 The bacterial cell and drug targets.

- *Inhibition of bacterial cell wall synthesis* leads to bacterial cell lysis and death. Agents operating in this way include penicillins, cephalosporins, and glycopeptides such as vancomycin. As animal cells do not have a cell wall, they are unaffected by such agents.
- *Interactions with the plasma membrane*: some antibacterial agents interact with the plasma membrane of bacterial cells to affect membrane permeability. This has fatal results for the cell. Polymyxins and tyrothricin operate in this way.
- *Disruption of protein synthesis* means that essential proteins and enzymes required for the cell's survival can no longer be made. Agents which disrupt protein synthesis include the rifamycins, aminoglycosides, tetracyclines, and chloramphenicol.
- *Inhibition of nucleic acid transcription and replication* prevents cell division and/or the synthesis of essential proteins. Agents acting in this way include nalidixic acid and proflavine.

We now consider these mechanisms in more detail.

19.4 Antibacterial agents which act against cell metabolism (antimetabolites)

19.4.1 Sulphonamides

19.4.1.1 The history of sulphonamides

The best example of antibacterial agents acting as antimetabolites are the sulphonamides (sometimes called the sulpha drugs). The sulphonamide story began in 1935 when it was discovered that a red dye called **prontosil** (Fig. 19.3) had antibacterial properties *in vivo* (i.e. when given to laboratory animals). Strangely enough, no antibacterial effect was observed *in vitro*. In other words, prontosil could not kill bacteria grown in the test tube. This remained a mystery until it was discovered that prontosil was metabolized by bacteria present in the small intestine of the test animal to give a product called **sulphanilamide** (Fig. 19.3). It was this compound which was the true antibacterial agent. Thus, prontosil was an early example of a **prodrug** (section 14.6). Sulphanilamide was synthesized in the laboratory and became the first synthetic antibacterial agent found to be active against a wide range of infections. Further developments led to a range of sulphonamides which proved effective against Gram-positive organisms, especially pneumococci and meningococci.

Despite their undoubted benefits, sulpha drugs have proved ineffective against infections such as *Salmonella*—the organism responsible for typhoid. Other problems have resulted from the way these drugs are metabolized, as toxic products are frequently obtained. This led to the sulphonamides being superseded by penicillin.

19.4.1.2 Structure–activity relationships

The synthesis of a large number of sulphonamide analogues (Fig. 19.4) led to the following conclusions:

- the *para*-amino group is essential for activity and must be unsubstituted (i.e. $R^1 = H$). The only exception is when R^1 = acyl (i.e. amides). The amides themselves are inactive but can be metabolized in the body to regenerate the active compound (Fig. 19.5). Thus, amides can be used as sulphonamide prodrugs;
- the aromatic ring and the sulphonamide functional group are both required;
- both the sulphonamide and amino group must be directly attached to the aromatic ring;
- the aromatic ring must be *para*-substituted only. Extra substitution eliminates activity for steric reasons;
- the sulphonamide nitrogen must be primary or secondary;
- R^2 is the only possible site that can be varied in sulphonamides.

19.4.1.3 Sulphanilamide analogues

In sulphanilamide analogues (Fig. 19.4), R^2 is often varied by incorporating a large range of heterocyclic or aromatic

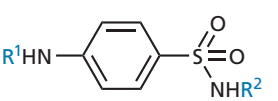


FIGURE 19.4 Sulphonamide analogues used in structure–activity relationship studies.

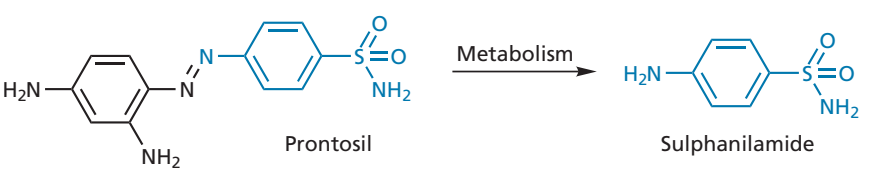


FIGURE 19.3 Metabolism of prontosil.

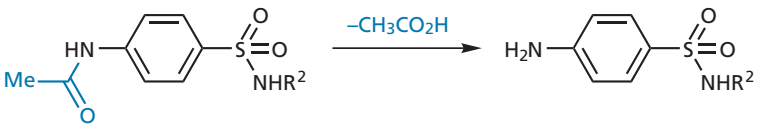


FIGURE 19.5 Metabolism of an *N*-acyl group to regenerate an active sulphonamide.

structures which affects the extent to which the drug binds to plasma protein. This, in turn, controls the blood levels and lifetime of the drug. Thus, a drug that binds strongly to plasma protein will be released slowly into the blood circulation and will be longer lasting. Varying R^2 can also affect the solubility of sulphonamides. To conclude, variations of R^2 affect the pharmacokinetics of the drug, rather than its mechanism of action (Box 19.1).

19.4.1.4 Applications of sulphonamides

Before the appearance of penicillin, the sulpha drugs were the drugs of choice in the treatment of infectious diseases. Indeed, they played a significant part in world history by saving Winston Churchill's life during World War II. After visiting North Africa for the Casablanca conference in 1943, Churchill became gravely ill with

BOX 19.1 Sulphonamide analogues with reduced toxicity

The primary amino group of sulphonamides is acetylated in the body and the resulting amides have reduced solubility which can lead to toxic effects. For example, the metabolite formed from **sulphathiazole** (an early sulphonamide) is poorly soluble and can prove fatal if it blocks the kidney tubules (Fig. 1). It is interesting to note that certain populations are more susceptible to this than others. For example, the Japanese and Chinese metabolize sulphathiazole more quickly than the average American and are more susceptible to its toxic effects.

It was discovered that the solubility problem could be overcome by replacing the thiazole ring in sulphathiazole with a pyrimidine ring to give **sulphadiazine** (Fig. 2). The reason for the improved solubility lies in the acidity of the sulphonamide NH proton. In sulphathiazole, this proton is not very acidic (high pK_a). Therefore, sulphathiazole and its metabolite are mostly un-ionized at blood pH. Replacing the thiazole ring with a more electron-withdrawing pyrimidine ring increases the acidity of the NH proton by stabilizing the resulting anion. Therefore, sulphadiazine and its metabolite are significantly ionized at blood pH. As a consequence, they are more soluble and less toxic. Sulphadiazine was also found to be more active than sulphathiazole and soon replaced it in therapy. Silver sulphadiazine cream is still used topically to prevent infection of burns, although it is really the silver ions which provide the antibacterial effect.

FIGURE 1 Metabolism of sulphathiazole.

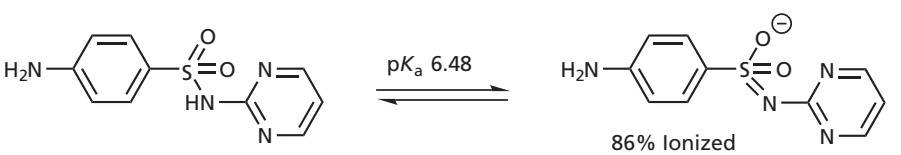


FIGURE 2 Sulphadiazine.

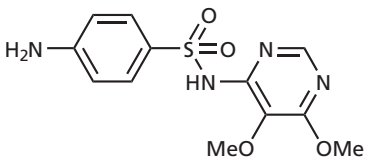


FIGURE 19.6 Sulphadoxine.

an infection and was bedridden for several weeks. Fortunately, he responded to the novel sulphonamide drugs of the day.

Penicillins largely superseded sulphonamides and, for a long time, sulphonamides took a back seat. There has been a revival of interest, however, with the discovery of a new 'breed' of longer-lasting sulphonamides. One example of this new generation is **sulphadoxine** (Fig. 19.6), which is so stable in the body that it need only be taken once a week. The combination of sulphadoxine and **pyrimethamine** is called **Fansidar** and has been used for the treatment of malaria.

The sulpha drugs presently have the following applications in medicine:

- treatment of urinary tract infections;
- eye lotions;
- treatment of infections of mucous membranes;
- treatment of gut infections (Box 19.2).

It is also worth noting that sulphonamides have occasionally found uses in other areas of medicine (section 12.4.4.2).

19.4.1.5 Mechanism of action

The sulphonamides act as competitive enzyme inhibitors of **dihydropteroate synthetase** and block the biosynthesis of **tetrahydrofolate** in bacterial cells (Fig. 19.7). Tetrahydrofolate is important in both human and bacterial cells, because it is an enzyme cofactor that provides one carbon units for the synthesis of the pyrimidine nucleic acid bases required for DNA synthesis (section 21.3.1). If pyrimidine and DNA synthesis is blocked, then the cell can no longer grow and divide.

Note that sulphonamides do not actively kill bacterial cells. They do, however, prevent the cells growing and multiplying. This gives the body's own defence systems enough time to gather their resources and wipe out the invader. Antibacterial agents which inhibit cell growth are classed as **bacteriostatic**, whereas agents such as penicillin which actively kill bacterial cells are classed as

BOX 19.2 Treatment of intestinal infections

Sulphonamides have been particularly useful against intestinal infections, and can be targeted against these by the use of prodrugs. For example, **succinyl sulphathiazole** is a prodrug of sulphathiazole (Fig. 1). The succinyl moiety contains an acidic group which means that the prodrug is ionized in the intestine. As a result, it is not absorbed into the bloodstream and is retained in the intestine. Slow enzymatic hydrolysis of the succinyl group then releases the active sulphathiazole where it is needed.

Benzoyl substitution (Fig. 2) on the aniline nitrogen has also given useful prodrugs that are poorly absorbed through the gut wall because they are too hydrophobic (section 11.3). They can be used in the same way.

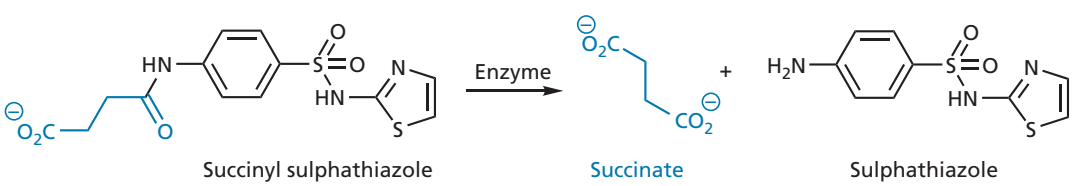


FIGURE 1 Succinyl sulphathiazole is a prodrug of sulphathiazole.

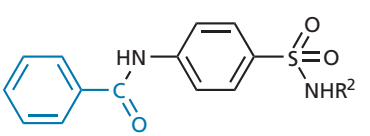


FIGURE 2 Substitution on the aniline nitrogen with benzoyl groups.

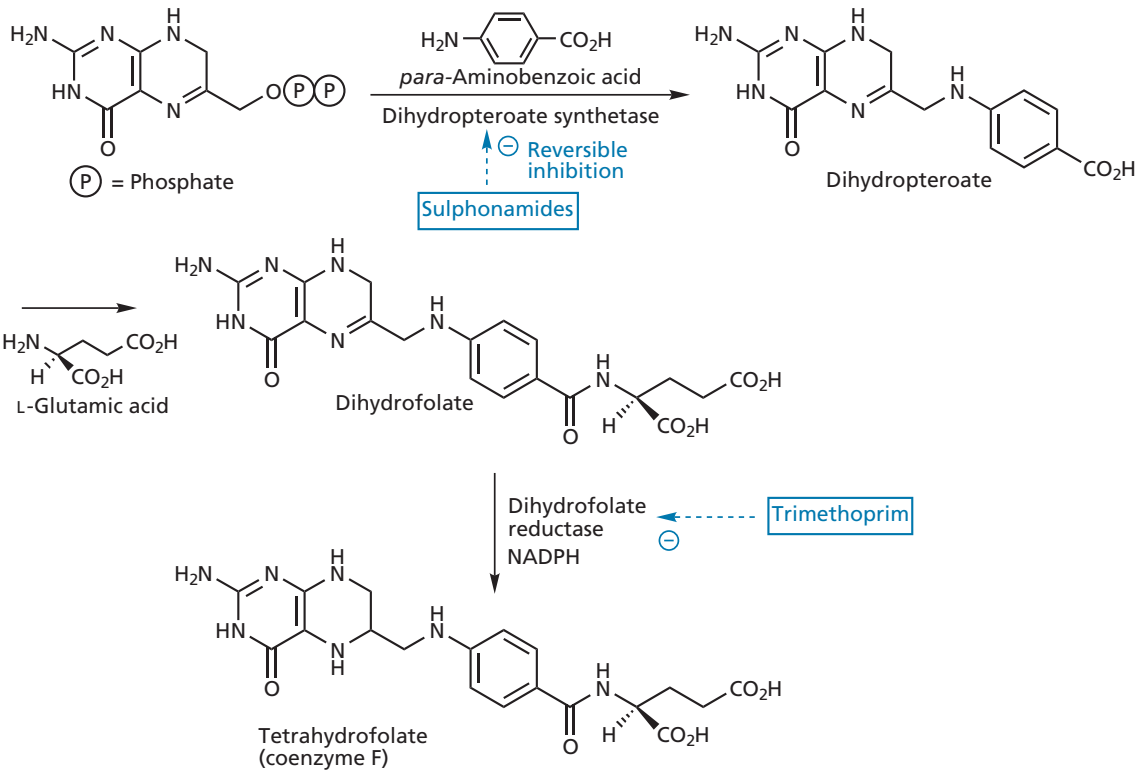


FIGURE 19.7 Mechanism of action of sulphonamides.

bactericidal. Because sulphonamides rely on a healthy immune system to complete the job they have started, they are not recommended for patients with a weakened immune system. This includes people with AIDS, as well as patients who are undergoing cancer chemotherapy or have had an organ transplant and are taking immunosuppressant drugs.

Sulphonamides act as inhibitors by mimicking ***p*-aminobenzoic acid** (PABA)—one of the normal substrates for dihydropteroate synthetase. The sulphonamide molecule is similar enough in structure to PABA that the enzyme is fooled into accepting it into its active site (Fig. 19.8). Once it is bound, the sulphonamide prevents PABA from binding. As a result, dihydropteroate is no longer synthesized. One might ask why the enzyme does not join the sulphonamide to the other component of dihydropteroate to give a dihydropteroate analogue containing the sulphonamide skeleton. This can in fact occur, but it does the cell no good at all because the analogue is not accepted by the next enzyme in the biosynthetic pathway.

Sulphonamides are competitive enzyme inhibitors so inhibition is reversible. This is demonstrated by certain organisms, such as staphylococci, pneumococci, and gonococci, which can acquire resistance by synthesizing more PABA. The more PABA there is in the cell, the more effectively it can compete with the sulphonamide inhibitor to reach the enzyme's active site. In such cases, the dose levels of sulphonamide have to be increased to bring

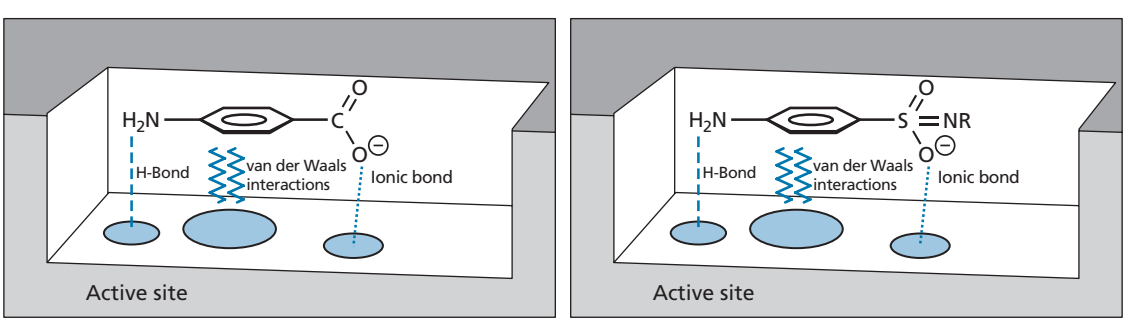


FIGURE 19.8 Sulphonamide prevents PABA from binding by mimicking PABA.

back the same level of inhibition. Resistance to sulphonamides can also arise by mutations that modify the target enzyme such that it has less affinity for sulphonamides, or by decreased permeability of the cell membrane to the sulphonamide.

Tetrahydrofolate is clearly necessary for the survival of bacterial cells, but it is also vital for the survival of human cells, so why are the sulpha drugs not toxic to humans? The answer lies in the fact that human cells synthesize tetrahydrofolate in a different manner and do not contain the enzyme dihydropteroate synthetase. In human cells, tetrahydrofolate is synthesized from **folic acid**, which is obtained from the diet as a vitamin and is brought across cell membranes by a transport protein.

We could now ask 'If human cells can acquire folic acid from the diet, why can't bacterial cells infecting the human body do the same, then convert it to tetrahydrofolate?' In fact, bacterial cells are unable to acquire folic acid because they lack the necessary transport protein required to carry it across the cell membrane.

To sum up, the success of sulphonamides is due to two metabolic differences between mammalian and bacterial cells:

- bacteria have a susceptible enzyme which is not present in mammalian cells;
- bacteria lack the transport protein that would allow them to acquire folic acid from outside the cell.

19.4.2 Examples of other antimetabolites

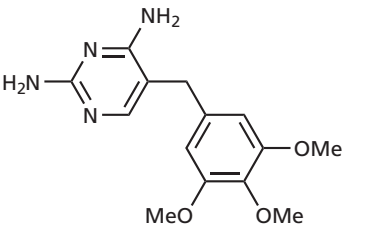
Other antimetabolites in medical use include **trimethoprim** and a group of compounds known as **sulphones** (Fig. 19.9).

19.4.2.1 Trimethoprim

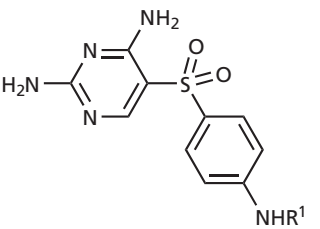
Trimethoprim is an orally active diaminopyrimidine structure, which has proved to be a highly selective antibacterial and antimalarial agent. It acts against **dihydrofolate reductase**—the enzyme which carries out the conversion of dihydrofolate to tetrahydrofolate—leading to the inhibition of DNA synthesis and cell growth.

Dihydrofolate reductase is present in mammalian cells, as well as bacterial cells, but mutations over millions of years have resulted in a significant difference in structure between the two enzymes such that trimethoprim recognizes and inhibits the bacterial enzyme more strongly. In fact, trimethoprim is 100,000 times more active against the bacterial enzyme.

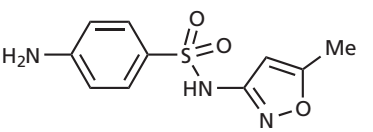
Trimethoprim is often given in conjunction with the sulphonamide **sulphamethoxazole** (Fig. 19.9) in a preparation called **cotrimoxazole**. The sulphonamide inhibits the incorporation of PABA into dihydropteroate,



Trimethoprim (antimalarial)



Sulphones (anti-leprosy)



Sulphamethoxazole

FIGURE 19.9 Examples of antimetabolites in medical use.

while trimethoprim inhibits dihydrofolate reductase. Therefore, two enzymes in the one biosynthetic route are inhibited (Fig. 19.7). This is a very effective method of inhibiting a biosynthetic route and has the advantage that the doses of both drugs can be kept down to a safe level. To get the same level of inhibition using a single drug, the dose level would have to be much higher, leading to possible side effects. This approach has been described as **sequential blocking**.

Resistance to trimethoprim has been observed in strains of *E. coli* which produce a new form of the target enzyme that has less affinity for the drug.

19.4.2.2 Sulphones

The sulphones (Fig. 19.9) are the most important drugs used in the treatment of leprosy. It is believed that they inhibit the same bacterial enzyme inhibited by the sulphonamides (i.e. dihydropteroate synthetase).

KEY POINTS

- The principle of chemotherapy or the magic bullet involves the design of chemicals which show selective toxicity against bacterial cells rather than mammalian cells.
- Early antibacterial agents were salvarsan, prontosil, and the sulphonamides. Following the discovery of penicillin, several classes of antibiotics were isolated from fungal strains.

- The bacterial cell differs in various respects from mammalian cells, allowing the identification of drug targets which are unique to bacterial cells, or which differ significantly from equivalent targets in mammalian cells.
- Antibacterial agents act on five main targets—cell metabolism, the cell wall, the plasma membrane, protein synthesis, and nucleic acid function.
- Sulphonamides require a primary aromatic amine group and a secondary sulphonamide group for good activity.
- Adding an aromatic or heteroaromatic group to the sulphonamide nitrogen provides a variety of sulphonamides with different pharmacokinetic properties.
- *N*-Acetylation of sulphonamides is a common metabolic reaction.
- Sulphonamides are used to treat infections of the urinary tract, gastrointestinal tract, and mucous membranes. They are also used in eye lotions.
- Sulphonamides are similar in structure to *para*-aminobenzoic acid—a component of dihydropteroate. As a result, they can bind to the bacterial enzyme responsible for dihydropteroate synthesis and act as an inhibitor.
- Mammals synthesize tetrahydrofolate from folic acid acquired from the diet. They lack the enzyme targeted by sulphonamides. Bacteria lack the transport mechanisms required to transport folic acid into their cells.
- Trimethoprim inhibits dihydrofolate reductase—an enzyme which converts folic acid to tetrahydrofolate. It has been used in combination with sulphamethoxazole in a strategy known as sequential blocking.
- Sulphones are used in the treatment of leprosy.

19.5 Antibacterial agents which inhibit cell wall synthesis

19.5.1 Penicillins

19.5.1.1 History of penicillins

In 1877, Pasteur and Joubert discovered that certain moulds produced toxic substances which killed bacteria. Unfortunately, these substances were also toxic to humans and were of no clinical value. They did demonstrate, however, that moulds could be a potential source of antibacterial agents.

In 1928, Fleming noted that a bacterial culture that had been left several weeks open to the air had become infected by a fungal colony. Of more interest was the fact that there was an area surrounding the fungal colony where the bacterial colonies were dying. He correctly concluded that the fungal colony was producing an antibacterial agent which was spreading into the surrounding area. Recognizing the significance of this, he set out to culture and identify the fungus, and showed it to be a relatively rare species of *Penicillium*. It has since been suggested that the *Penicillium* spore responsible for the fungal colony originated from another laboratory in the building, and that the spore was carried by air currents to be blown through the window of Fleming's laboratory. This in itself appears to be a remarkable stroke of good fortune. However, a series of other chance events were involved in the story—not least the weather! A period of early cold weather had encouraged the fungus to grow while the bacterial colonies had remained static. A period of warm weather then followed which encouraged the bacteria to grow. These weather conditions were the ideal experimental conditions required for:

- the fungus to produce penicillin during the cold spell;
- the antibacterial properties of penicillin to be revealed during the hot spell.

If the weather had been consistently cold, the bacteria would not have grown significantly and the death of cell colonies close to the fungus would not have been seen. Alternatively, if the weather had been consistently warm, the bacteria would have outgrown the fungus and little penicillin would have been produced. As a final twist to the story, the crucial agar plate had been stacked in a bowl of disinfectant ready for washing up, but was actually placed above the surface of the disinfectant. It says much for Fleming's observational powers that he bothered to take any notice of a discarded culture plate and that he spotted the crucial area of inhibition.

Fleming spent several years investigating the novel antibacterial extract and showed it to have significant antibacterial properties while being remarkably nontoxic to mammals. Unfortunately, Fleming was unable to isolate and purify the active principle, and he came to the conclusion that penicillin was too unstable to be used clinically.

The problem of isolating penicillin was eventually solved in 1938 by Florey and Chain by using processes such as freeze-drying and chromatography, which allowed isolation of the antibiotic under much milder conditions than had previously been available. By 1941, Florey and Chain were able to carry out the first clinical trials on crude extracts of penicillin and achieved spectacular success. Further developments aimed at producing the new agent in large quantities were developed in the USA, and, by 1944, there was enough penicillin to treat casualties arising from the D-Day landings.

Although the use of penicillin was now widespread, the structure of the compound was still not settled and the unusual structures being proposed proved a source

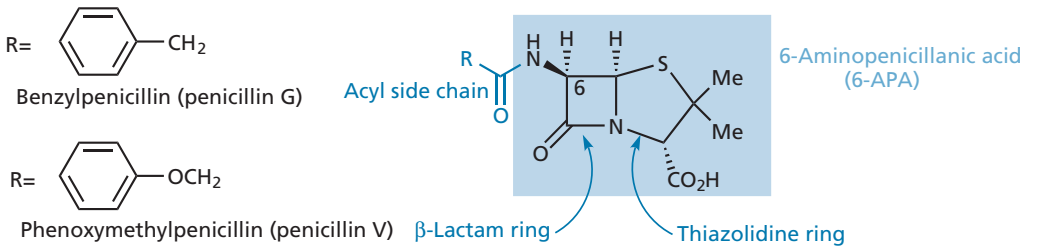


FIGURE 19.10 The structure of penicillin.

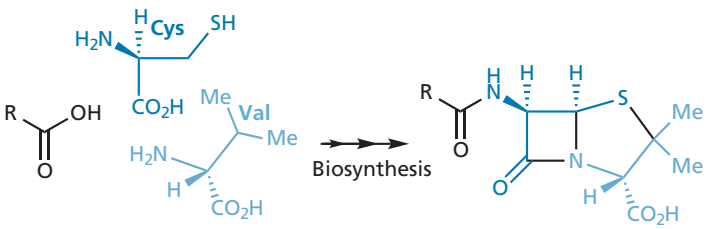


FIGURE 19.11 The biosynthetic precursors of penicillin.

of furious debate. The issue was finally settled in 1945 when Dorothy Hodgkins established the structure by X-ray crystallographic analysis. The structure was quite surprising at the time, as penicillin was clearly a highly strained molecule, which explained why Fleming had been unsuccessful in purifying it.

The full synthesis of such a highly strained molecule presented a huge challenge—one that was met successfully by Sheehan in 1957. Unfortunately, the full synthesis was too involved to be of commercial use, but, in the following year, Beechams isolated a biosynthetic intermediate of penicillin called **6-aminopenicillanic acid (6-APA)**. This revolutionized the field of penicillins by providing the starting material for a huge range of **semi-synthetic** penicillins.

Since then, penicillins have been used widely and often carelessly. As a result, penicillin-resistant bacteria have evolved and have become an increasing problem. The fight against penicillin-resistant bacteria was helped in 1976 when Beechams discovered a natural product called **clavulanic acid**, which proved highly effective in protecting penicillins from the bacterial enzymes which attack them (section 19.5.4.1).

19.5.1.2 Structure of benzylpenicillin and phenoxymethylpenicillin

Penicillin (Fig. 19.10) contains a highly unstable looking bicyclic system consisting of a four-membered β-lactam ring fused to a five-membered thiazolidine ring. The skeleton of the molecule suggests that it is derived from the amino acids cysteine and valine, and this has been established (Fig. 19.11). The overall shape of the molecule is like a half-open book, as shown in Fig. 19.12.

The acyl side chain (R) varies, depending on the components of the fermentation medium. For example, corn steep liquor (the fermentation medium first used for mass production of penicillin) contains high levels of phenylacetic acid ($PhCH_2CO_2H$) and gives **benzylpenicillin (penicillin G**; R = benzyl). A fermentation medium containing phenoxyacetic acid ($PhOCH_2CO_2H$) gives **phenoxymethylpenicillin** (**penicillin V**; R = $PhOCH_2$) (Fig. 19.10).

Test your understanding and practise your molecular modelling with Exercise 19.1.

19.5.1.3 Properties of benzylpenicillin

Benzylpenicillin (penicillin G) is active against a range of bacterial infections (Box 19.3) and lacks serious side effects for most patients. However, there are various drawbacks. It cannot be taken orally because it is broken down by stomach acids, it has a narrow spectrum of activity, and there are many bacterial infections against which it has no effect—particularly those where the microorganism produces an enzyme called β-**lactamase**. This is an enzyme which hydrolyses the β-lactam ring of benzylpenicillin and makes

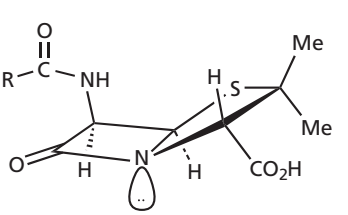


FIGURE 19.12 The three-dimensional shape of penicillin.

BOX 19.3 Clinical properties of benzylpenicillin and phenoxymethylpenicillin

Benzylpenicillin is active against non-β-lactamase-producing Gram-positive bacilli (e.g. *Meningitis*, *Gonorrhoea,* and early strains of staphylococci) and several Gram-negative cocci (e.g. *Neisseria*). It is effective for many streptococcal, pneumococcal, gonococcal, and meningococcal infections. It is also used to treat anthrax, diphtheria, gas-gangrene, leptospirosis, and Lyme disease in children. It can be effective against tetanus, although **metronidazole** is preferred. As penicillin is bactericidal, it is most active against rapidly dividing bacteria. There are many bacterial species against which benzyl penicillin shows no activity, in particular Gram-negative bacteria and those producing β-lactamase enzymes. It is ineffective when taken orally and should be administered by intravenous or intramuscular injection.

Phenoxymethylpenicillin is recommended for the treatment of various problems such as tonsillitis, rheumatic fever, otitis media, and oral infections.

Allergic reactions are suffered by some individuals when they take penicillins, varying from a rash to immediate anaphylactic shock. **Anaphylactic reactions** occur in 0.2% of patients with a fatality rate of 0.001%. Less serious allergic reactions are more common (1–4%).

it inactive. Therefore, there is scope for producing analogues with improved properties. Before looking at penicillin analogues, we shall look at penicillin's mechanism of action.

19.5.1.4 Mechanism of action for penicillin

Structure of the cell wall

In order to understand penicillin's mechanism of action, we have to first look at the structure of the bacterial cell wall and the mechanism by which it is formed. Bacteria have cell walls in order to survive a large range of environmental conditions, such as varying pH, temperature, and osmotic pressure. Without a cell wall, water would continually enter the cell as a result of osmotic pressure, causing the cell to swell and burst (lysis). The cell wall is very porous and does not block the entry of water, but it does prevent the cell swelling. Animal cells do not have a cell wall, making it the perfect target for antibacterial agents such as penicillins.

The wall is a peptidoglycan structure (Fig. 19.13). In other words, it is made up of peptide and sugar units. The structure of the wall consists of a parallel series of sugar backbones containing two types of sugar [*N*-**acetylmuramic acid** (**NAM**) and *N*-**acetylglucosamine** (**NAG**)] (Fig. 19.14). Peptide chains are bound to the NAM sugars and it is interesting to note the presence of D-amino acids in these chains. In human biochemistry there are only L-amino acids, whereas bacteria have **racemase** enzymes that can convert L-amino acids into D-amino acids. In the

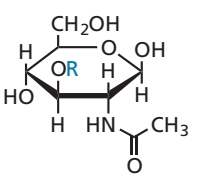


FIGURE 19.14 Sugars contained in the cell wall structure of bacteria. R = H, *N*-acetylglucosamine (NAG); R = $CHMeCO_2H$, *N*-acetylmuramic acid (NAM).

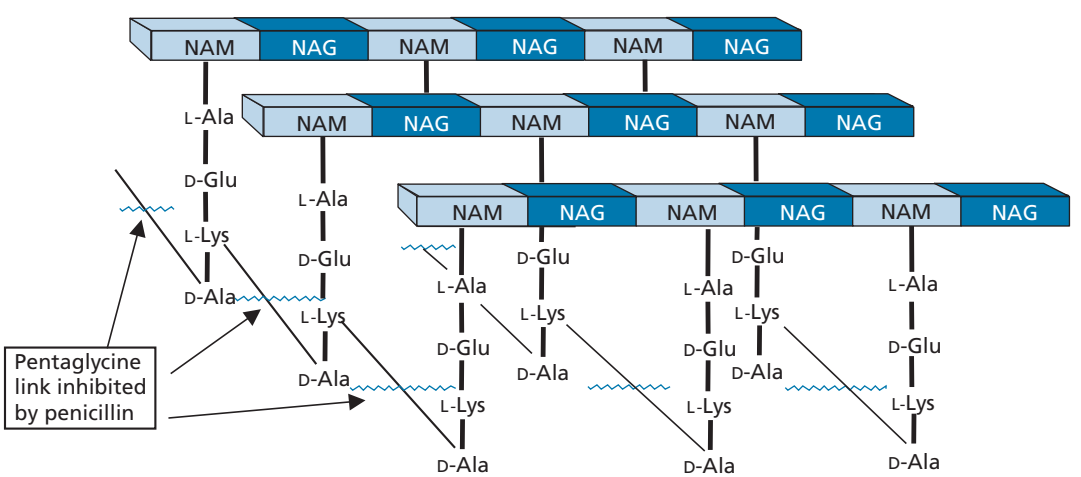


FIGURE 19.13 Peptidoglycan structure of bacterial cell walls.

final stage of cell wall biosynthesis, the peptide chains are linked together by the displacement of D-alanine from one chain by glycine in another.

About 30 enzymes are involved in the overall biosynthesis of the cell wall, but it is the final cross-linking reaction which is inhibited by penicillin. This leads to a cell wall framework that is no longer interlinked (Fig. 19.15). As a result, the wall becomes fragile and can no longer prevent the cell from swelling and bursting. The enzyme responsible for the cross-linking reaction is known as the **transpeptidase enzyme**. There are several types of the enzyme which vary in character from one bacterial species to another, but they are all inhibited to various degrees by penicillins.

There are significant differences in the thickness of the cell wall between Gram-positive and Gram-negative bacteria. The cell wall in Gram-positive bacteria consists of 50–100 peptidoglycan layers, whereas in Gram-negative bacteria it consists of only two layers.

The transpeptidase enzyme and its inhibition

The transpeptidase enzyme is bound to the outer surface of the cell membrane and is similar to a class of enzymes called the **serine proteases**, so called because they contain a serine residue in the active site and catalyse the hydrolysis of peptide bonds. In the normal mechanism (see Fig. 19.16a), serine acts as a nucleophile to split the peptide bond between the two unusual D-alanine units on a peptide chain. The terminal alanine departs the active site, leaving the peptide chain bound to the active site. The pentaglycyl moiety of another peptide chain now enters the active site and the terminal glycine forms a peptide bond to the alanine group, displacing it from serine and linking the two chains together.

It has been proposed that penicillin has a conformation which is similar to the transition-state conformation taken up by the D-Ala-D-Ala moiety during the cross-linking reaction, and that the enzyme mistakes

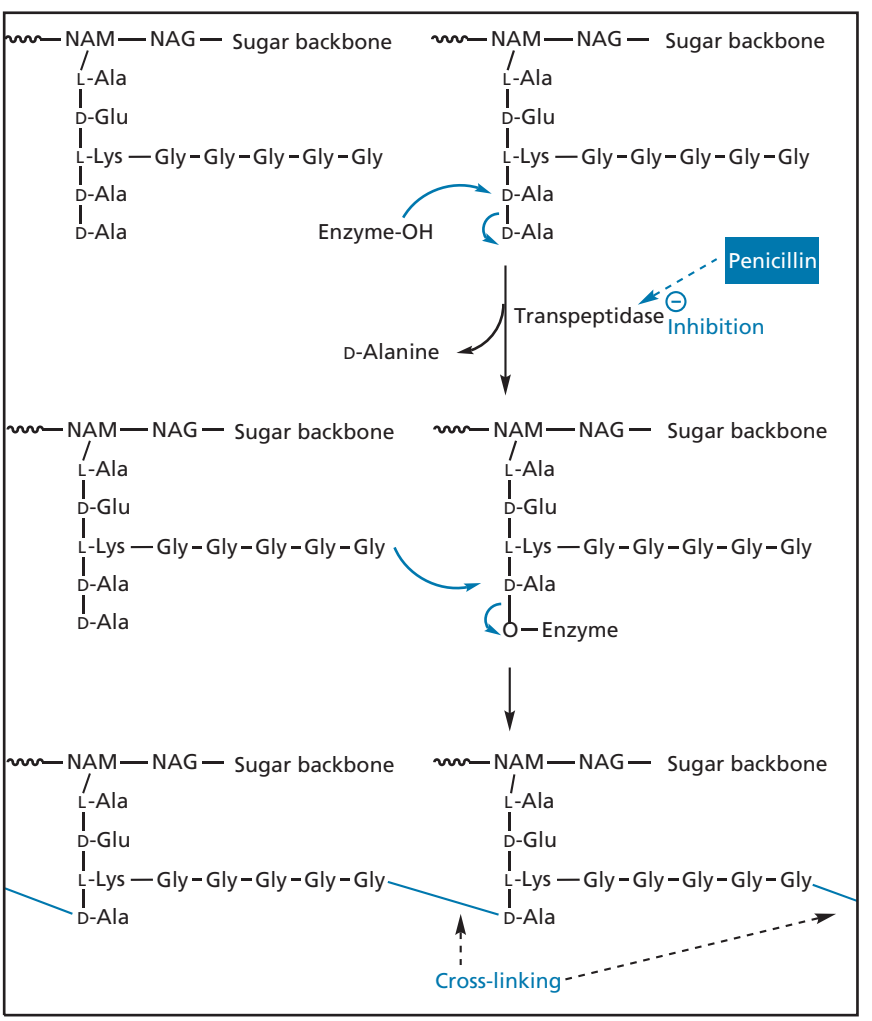


FIGURE 19.15 Cross-linking of bacterial cell walls inhibited by penicillin.

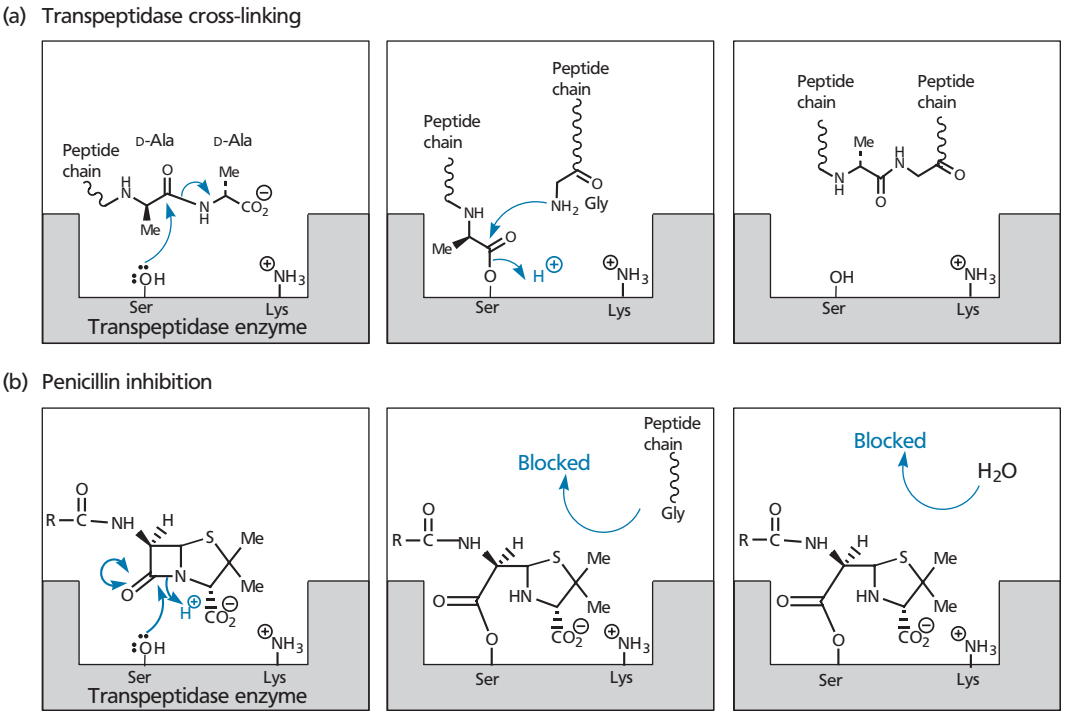


FIGURE 19.16 Mechanisms of transpeptidase cross-linking and penicillin inhibition.

penicillin for D-Ala-D-Ala and binds it to the active site. Once bound, penicillin is subjected to nucleophilic attack by serine (Fig. 19.16).

The enzyme can attack the β-lactam ring of penicillin and cleave it in the same way as it did with the peptide bond. However, penicillin is cyclic so the molecule is not split in two and nothing leaves the active site. Subsequent hydrolysis of the ester group linking the penicillin to the active site does not take place either, as the penicillin structure blocks access to the pentaglycine chain or water.

If penicillin *is* acting as a mimic for a D-Ala-D-Ala moiety, this provides another explanation for its lack of toxicity. Since there are no D-amino acids or D-Ala-D-Ala segments in any human protein, it is unlikely that any of the body's serine protease enzymes would recognize either the segment or penicillin itself. As a result, penicillin is selective for the bacterial transpeptidase enzyme and is ignored by the body's own serine proteases.

This theory has one or two anomalies, though. For example, **6-methylpenicillin** (Fig. 19.17) was thought to be a closer analogue to D-Ala-D-Ala. On that basis, it should fit the active site better and have higher activity. However, when this structure was synthesized, it was found to be inactive. It is now proposed that 6-methoxypenicillin is a closer analogue to acyl-D-Ala-D-Ala than 6-methylpenicillin. Indeed, antibacterial penicillin structures containing a 6-methoxy substituent have been developed, for example **temocillin** (Fig. 19.27). Molecular modelling studies involving overlays of penicillin analogues (section 17.9) have demonstrated that the methyl group of a 6-methoxy substituent is more closely aligned to the methyl group of acyl-D-Ala-D-Ala, than a 6-methyl group would be (see Molecular modelling exercise 19.2).

19.5.1.5 Resistance to penicillin

Bacterial strains vary in their susceptibility to penicillin. Some species, such as streptococci, are quite vulnerable, whereas a bacterium like *Pseudomonas aeruginosa* is particularly resistant (see Box 19.4). Other species, such as

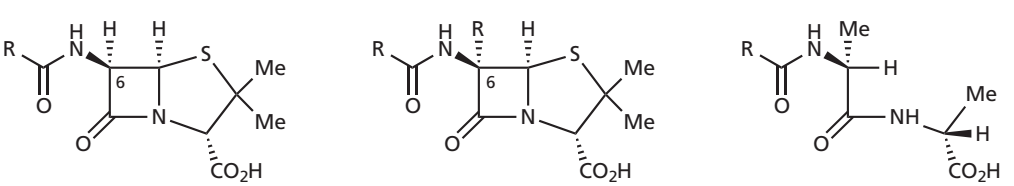


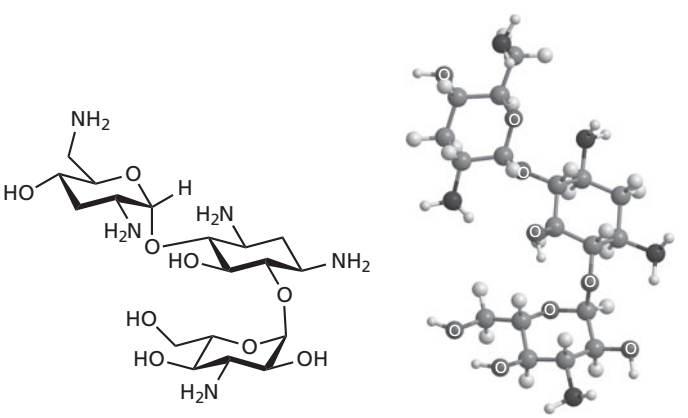
FIGURE 19.17 Comparison of penicillin, 6-substituted penicillins, and acyl-D-Ala-D-Ala.

BOX 19.4 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is an example of an **opportunistic pathogen**. Such organisms are not normally harmful to healthy individuals. Indeed, many people carry the organism without being aware of it, because their immune system keeps it under control. Once that immune system is weakened, though, the organism can start multiplying and lead to serious illness. Hospital-bound patients are particularly at risk, especially those suffering from shock or AIDS, or those undergoing cancer chemotherapy. Burn victims are particularly prone to *P. aeruginosa* skin infections and this can lead to septicaemia, which can prove fatal. The organism is also responsible for serious lung infections among patients undergoing mechanical ventilation.

The cells of *P. aeruginosa* are rod-shaped and can appear blue or green in colour, which is why it was given the name aeruginosa. It prefers to grow in moist environments and has been isolated from soil, water, plants, animals, and humans. It can even grow in distilled water and contact lens solutions. In hospitals, there are several possible sources of infection, including respiratory equipment, sinks, uncooked vegetables, and flowers brought by visitors.

Pseudomonas aeruginosa is a difficult organism to treat because it has an intrinsic resistance to a wide variety of antibacterial agents, including many penicillins, cephalosporins, tetracyclines, quinolones, and chloramphenicol. There are two reasons for this. The outer membrane of the cell has a low permeability to drugs and even if a drug does enter the cell, there is an efflux system which can pump it back out again. Nevertheless, there are drugs which have proved effective against the organism—in particular aminoglycosides such as tobramycin or gentamicin, and penicillins such as ticarcillin. These are often given in combination with each other.



Tobramycin.

S. aureus are initially vulnerable, but acquire resistance when they are exposed to penicillin over a period of time. There are several reasons for this varied susceptibility.

Physical barriers

If penicillin is to inhibit the transpeptidase enzyme, it has to reach the outer surface of the bacterial cell membrane where the enzyme is located. Thus, penicillin has to pass through the cell walls of both Gram-positive and Gram-negative bacteria. The cell wall is much thicker in Gram-positive bacteria than in Gram-negative bacteria, so one might think that penicillin would be more effective against Gram-negative bacteria. However, this is not the case. Although the cell wall is a strong, rigid structure, it is also highly porous, which means that small molecules like penicillin can move through it without difficulty. One can imagine the cell wall being like several layers of chicken wire and the penicillin molecules as small pebbles able to pass through the gaps.

If the cell wall does not prevent penicillin reaching the cell membrane, what does? As far as Gram-positive bacteria are concerned there *is* no barrier and that is why penicillin G has good activity against these organisms. However, Gram-negative bacteria have an outer lipopolysaccharide membrane surrounding the cell wall which is impervious to water and polar molecules, such as penicillin (Fig. 19.18). That can explain why Gram-negative bacteria are generally resistant, but not why some Gram-negative bacteria are susceptible and some are not. Should they not all be resistant?

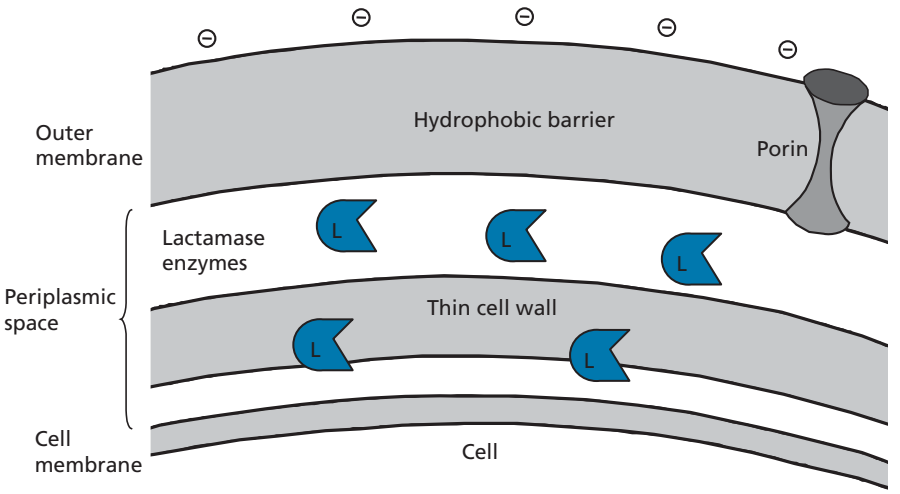


FIGURE 19.18 Outer surface of a Gram-negative bacterial cell.

The answer lies in protein structures called **porins**, which are located in the outer membrane. These act as pores through which water and essential nutrients can pass to reach the cell. Small drugs such as penicillin can also pass this way, but whether they do or not depends on the structure of the porin, as well as the characteristics of the penicillin (i.e. its size, structure, and charge). In general, drugs have less chance of passing through the porins if they are large, have a negative charge, and are hydrophobic. In contrast, a small hydrophilic drug that can exist as a zwitterion can pass through. Therefore, porins play a crucial role in controlling the amount of penicillin capable of reaching the periplasmic space between the outer membrane and cell membranes. If access is slow, the concentration of penicillin at the transpeptidase enzyme may be insufficient to inhibit it effectively.

Presence of β-lactamase enzymes

The presence of β-lactamase enzymes is the most important mechanism by which bacteria gain resistance to penicillin. β-lactamases are enzymes which have mutated from transpeptidases and so they are quite similar in nature. For example, they have a serine residue in the active site and can open up the β-lactam ring of penicillin to form an ester link to the structure. Unlike the transpeptidase enzyme, β-lactamases are able to hydrolyse the ester link and shed the ring-opened penicillin. They do this so effectively that 1000 penicillin molecules are hydrolysed per second (Fig. 19.19).

Some Gram-positive bacterial strains are resistant to penicillin because they can release β-lactamase into the surrounding environment such that penicillin is intercepted before it reaches the cell membrane. The enzyme eventually dissipates through the cell wall and is lost, so the bacterium has to keep generating the enzyme to maintain its protection. *Staphylococcus aureus* is a Gram-positive bacterium that used to be susceptible to penicillin, but 95% of *S. aureus* strains now release a β-lactamase which hydrolyses penicillin G.

Most, if not all, Gram-negative bacteria produce β-lactamases which makes them more resistant to penicillins. Moreover, the β-lactamase released is trapped in the periplasmic space between the cell membrane and the outer membrane because it cannot pass through the latter. As a result, any penicillin managing to penetrate the outer membrane encounters a higher concentration of β-lactamase than it would with Gram-positive bacteria. This might suggest, again, that all Gram-negative bacteria should be resistant to penicillin. However, there are various types of β-lactamase enzyme produced by both Gram-positive and Gram-negative bacteria, and

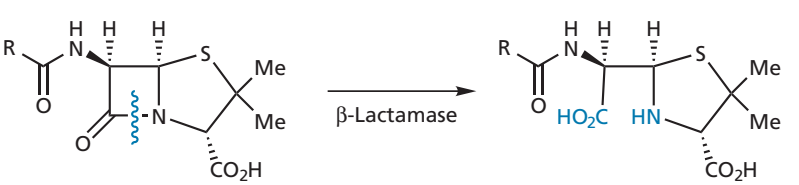


FIGURE 19.19 β-lactamase deactivation of penicillin.

these vary in their substrate selectivity. Some are selective for penicillins (**penicillinases**), some for cephalosporins (section 19.5.2) (**cephalosporinases**), and some for both penicillins and cephalosporins. The differing levels of enzyme and their differing affinities for different β-lactams account for the varying susceptibilities of Gram-negative bacteria to different β-lactams.

High levels of transpeptidase enzyme produced

In some Gram-negative bacteria, excess quantities of transpeptidase are produced and penicillin is incapable of inactivating all the enzyme molecules present.

Affinity of the transpeptidase enzyme to penicillin

There are several forms of the transpeptidase enzyme present within any bacterial cell and these vary in their affinity for the different β-lactams. Differences in the relative proportions of these enzymes across bacterial species account, in part, for the variable susceptibility of these bacteria to different penicillins. For example, early strains of *S. aureus* contained transpeptidase enzymes which had a high affinity for penicillin and were inhibited effectively. Penicillin-resistant strains of *S. aureus* acquired a transpeptidase enzyme called **penicillin binding protein 2a** (**PBP2a**), which has a much lower affinity to penicillins. The presence of low-affinity transpeptidases is also a problem with enterococci and pneumococci.

Transport back across the outer membrane of Gram-negative bacteria

There are proteins in the outer membrane of some Gram-negative bacteria which are capable of pumping penicillin out of the periplasmic space, thus lowering its concentration and effectiveness. The extent to which this happens varies from species to species and also depends on the structure of the penicillin. This is known as an **efflux** process.

Mutations and genetic transfers

Mutations can occur which will affect any or all of the above mechanisms such that they are more effective in resisting the effects of β-lactams. Small portions of DNA carrying the genes required for resistance can also be transferred from one cell to another by means of genetic vehicles called **plasmids**. These are small pieces of circular bacterial extra-chromosomal DNA. If the transferred DNA contains a gene coding for a β-lactamase enzyme or some other method of improved resistance, then the recipient cell acquires immunity. Genetic material can also be transferred between bacterial cells by viruses and by the uptake of free DNA released by dead bacteria.

19.5.1.6 Methods of synthesizing penicillin analogues

Having studied the mechanism of action of penicillin G and the various problems surrounding resistance, we now look at how analogues of penicillin G can be synthesized which might have improved stability and activity. A method of preparing analogues is required which is cheap, efficient, and flexible. Sheehan's full synthesis of penicillin is too long and low yielding (1%) to be practical, which limits the options to fermentation methods or semi-synthetic procedures.

Fermentation

Originally, the only way to prepare different penicillins was to vary the fermentation conditions. Adding different carboxylic acids to the fermentation medium resulted in penicillins with different acyl side chains (e.g. **phenoxymethylpenicillin**; Fig. 19.10). Unfortunately, there was a limitation to the sort of carboxylic acid which was accepted by the biosynthetic route (i.e. only acids of general formula RCH_2CO_2H). This, in turn, restricted the variety of analogues which could be obtained. The other major disadvantage was the tedious and time-consuming nature of the method.

Semi-synthetic procedure

In 1959, Beechams isolated a biosynthetic intermediate of penicillin from *Penicillium chrysogenum* grown in a fermentation medium which was deficient in a carboxylic acid. The intermediate (**6-aminopenicillanic acid; 6-APA**) proved to be one of Sheehan's synthetic intermediates, and so it was possible to use this to synthesize a huge number of analogues by a semi-synthetic method. Thus, fermentation yielded 6-APA, which could then be treated with a range of acid chlorides (Fig. 19.20).

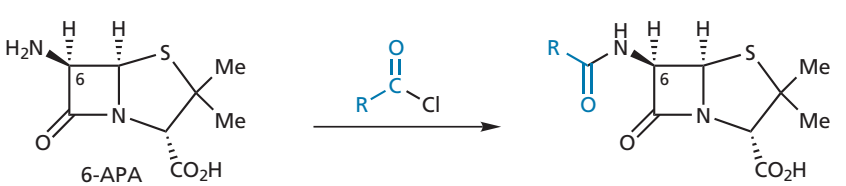


FIGURE 19.20 Penicillin analogues synthesized by acylating 6-APA.

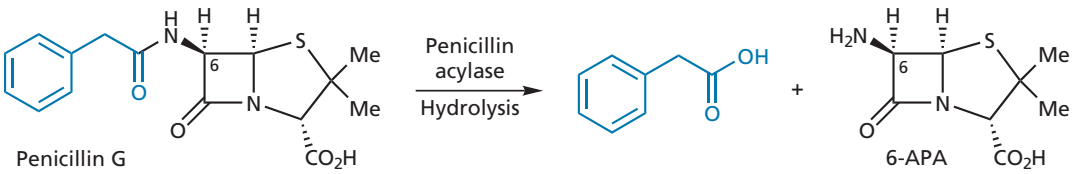


FIGURE 19.21 Synthesis of 6-APA from penicillin G.

6-APA is now produced more efficiently by hydrolysing penicillin G or penicillin V with an enzyme (**penicillin acylase**) (Fig. 19.21), or by a chemical method that allows the hydrolysis of the side chain in the presence of the highly strained β-lactam ring. The latter procedure is described in more detail in section 19.5.2.2, where it is used to hydrolyse the side chain from cephalosporins.

We have emphasized the drive to make penicillin analogues with varying acyl side chains, but what is so special about the acyl side chain? Could changes not be made elsewhere in the molecule? In order to answer these questions we need to look at the structure–activity relationships (SARs) of penicillins.

19.5.1.7 Structure–activity relationships of penicillins

A large number of penicillin analogues have been synthesized and studied. The results of these studies led to the following SAR conclusions (Fig. 19.22):

- the strained β-lactam ring is essential;
- the free carboxylic acid is essential. This is usually ionized and penicillins are administered as sodium or potassium salts. The carboxylate ion binds to the charged nitrogen of a lysine residue in the binding site;
- the bicyclic system is important. This confers further strain on the β-lactam ring—the greater the strain, the greater the activity, but the greater the instability of the molecule to other factors;
- the acylamino side chain is essential;
- sulphur is usual but not essential (see section 19.5.3);
- the stereochemistry of the bicyclic ring with respect to the acylamino side chain is important.

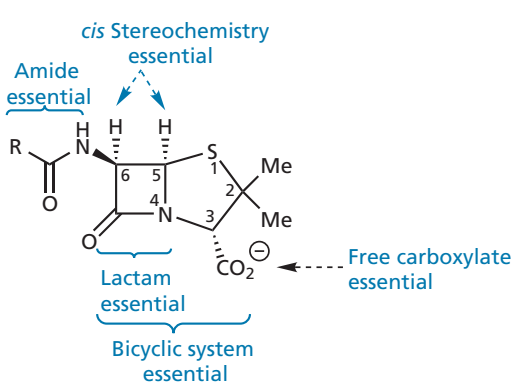


FIGURE 19.22 Structure–activity relationships of penicillins.

The results of this analysis led to the inevitable conclusion that very little variation is tolerated by the penicillin nucleus and that any variations are restricted to the acylamino side chain.

19.5.1.8 Penicillin analogues

In this section we consider the penicillin analogues which proved successful in tackling the problems of acid sensitivity, β-lactamase sensitivity, and limited breadth of activity.

Acid sensitivity of penicillins

There are three reasons for the acid sensitivity of penicillin G.

- *Ring strain*: the bicyclic system in penicillin consists of a four-membered ring fused to a five-membered ring. As a result, penicillin suffers large angle and torsional strains. Acid-catalysed ring-opening relieves these strains by breaking open the more highly strained β-lactam ring (Fig. 19.23).
- *A highly reactive β-lactam carbonyl group*: The carbonyl group in the β-lactam ring is highly susceptible to nucleophiles and does not behave like a normal tertiary amide. The latter is resistant to nucleophilic attack because the carbonyl group is stabilized by the neighbouring nitrogen atom, as shown in Fig. 19.24. The nitrogen can feed its lone pair of electrons into the carbonyl group to form a dipolar resonance structure with bond angles of 120°. This resonance stabilization is impossible for the β-lactam ring because of the increase in angle strain that would result in having a double bond within a four-membered β-lactam ring. The preferred bond angles for a double bond are 120° but the bond angles of the β-lactam ring are constrained to 90°. As a result, the lone pair is localized on the nitrogen atom and the carbonyl group is

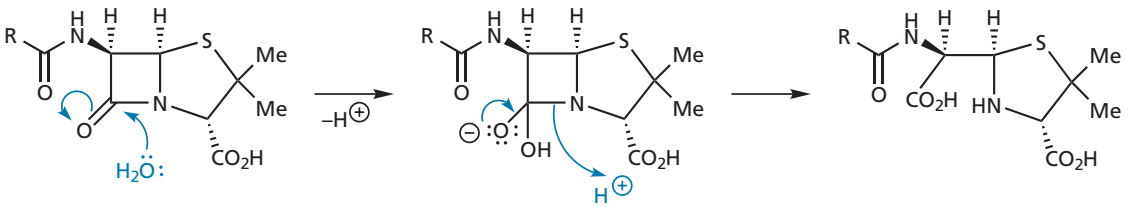


FIGURE 19.23 Ring-opening of the β-lactam ring under acidic conditions.

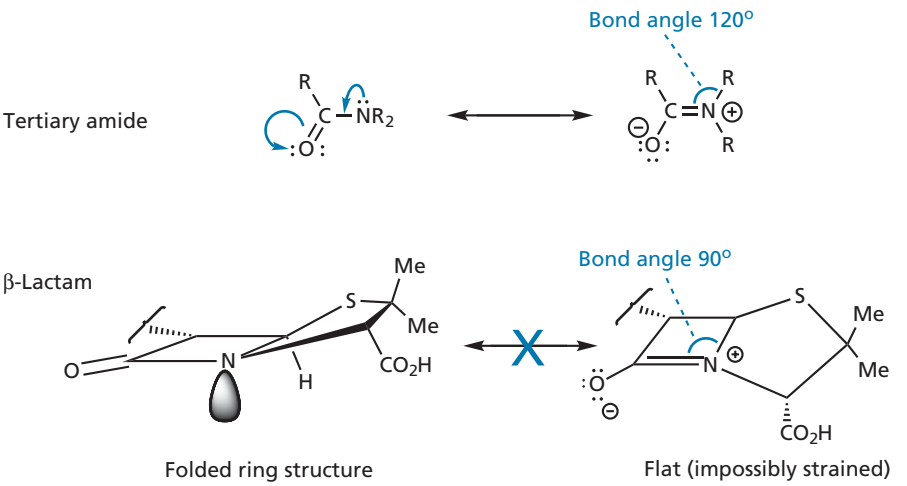


FIGURE 19.24 Comparison of *tertiary* amide and β-lactam carbonyl groups.

more electrophilic than one would expect for a tertiary amide.

- *Influence of the acyl side chain* (**neighbouring group participation**): Fig. 19.25 demonstrates how the neighbouring acyl group can actively participate in a mechanism to open up the lactam ring. Thus, penicillin G has a self-destruct mechanism built into its structure.

Acid-resistant penicillins

It can be seen that countering acid sensitivity is a difficult task. Nothing can be done about the first two factors, as the β-lactam ring is vital for antibacterial activity. Therefore, only the third factor can be tackled. The task then becomes one of reducing the amount of neighbouring group participation taking place. This was achieved by placing an electron-withdrawing group in the side chain which could draw electrons away from the carbonyl oxygen and reduce its tendency to act as a nucleophile (Fig. 19.26).

Phenoxymethylpenicillin (penicillin V) has an electronegative oxygen on the acyl side chain with the electron-withdrawing effect required. The molecule has better acid stability than penicillin G and is stable enough to survive the acid in the stomach, so it can be given orally.

Other penicillin analogues with an electron-withdrawing substituent (X) on the α-carbon of the side chain (Fig. 19.26) have also proved resistant to acid hydrolysis and can be given orally (e.g. **ampicillin**; see Fig. 19.29).

To conclude, the problem of acid sensitivity is fairly easily solved by having an electron-withdrawing group on the acyl side chain.

β-Lactamase-resistant penicillins

The problem of **β-lactamases** (or **penicillinases**) became critical in 1960, when the widespread use of penicillin G led to an alarming increase of penicillin-resistant *S. aureus* infections. At one point, 80% of all *S. aureus* infections in hospitals were due to virulent, penicillin-resistant strains. Alarmingly, these strains were also resistant to all other available antibiotics. Fortunately, a solution to the problem was just around the corner—the design of β-lactamase-resistant penicillins.

The strategy of steric shields (section 14.2.1) was used successfully to block penicillin from accessing the penicillinase or β-lactamase active site by placing a bulky group on the side chain (Fig. 19.27). However, there was a problem. If the steric shield was *too* bulky then it also prevented the penicillin from attacking the transpeptidase target enzyme. Therefore, a great deal of work had

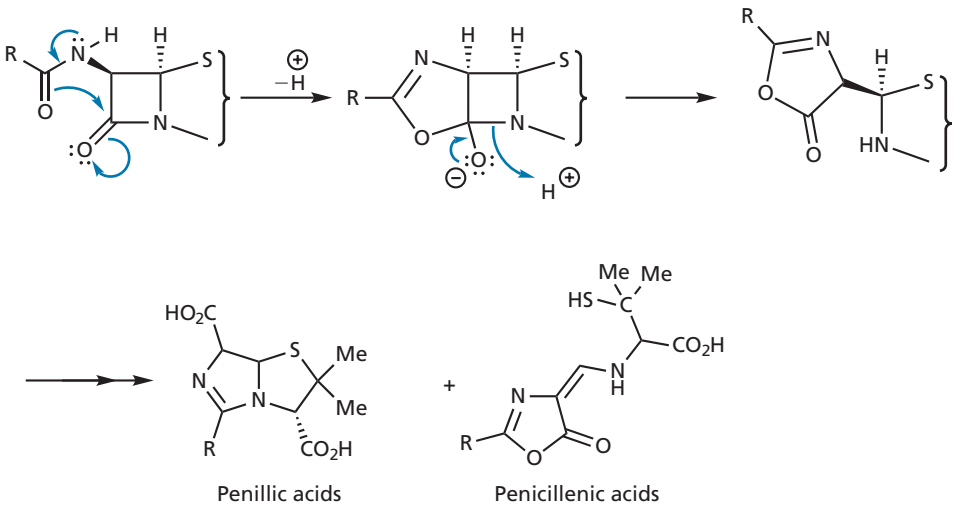


FIGURE 19.25 Influence of the acyl side chain on the acid sensitivity of penicillins.

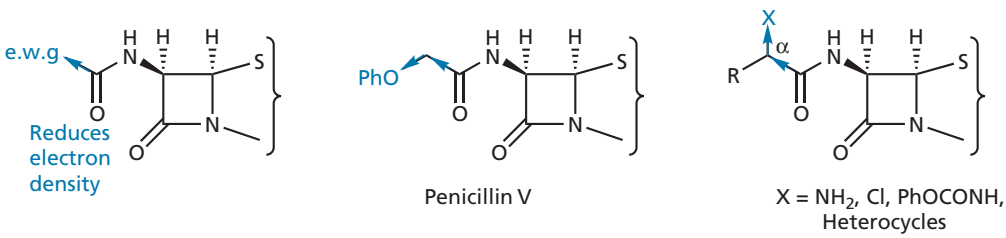


FIGURE 19.26 Reduction of neighbouring group participation with an electron-withdrawing group (e.w.g.).

to be done to find the ideal shield—one large enough to ward off the lactamase enzyme, but sufficiently small to allow the penicillin to bind to the target enzyme. The fact that the β-lactam ring interacts with both enzymes in the same way highlights the difficulty in achieving that goal.

Fortunately, shields *were* found which could make that discrimination. **Methicillin** (Fig. 19.27) was the first effective semi-synthetic penicillin with resistance to the *S. aureus* β-lactamase enzyme and reached the clinic just in time to treat the growing *S. aureus* problem. The

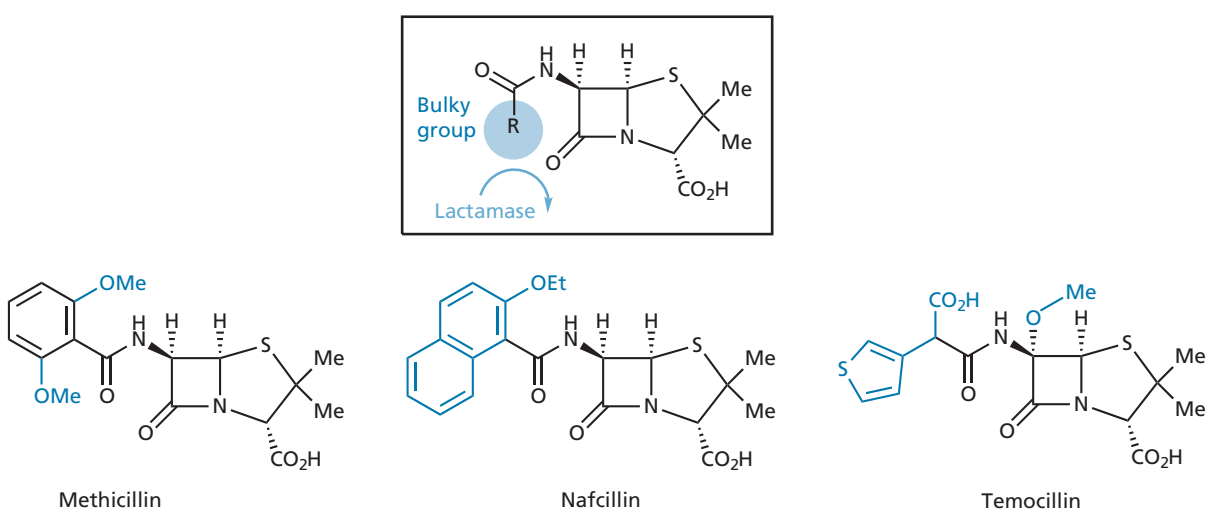


FIGURE 19.27 The use of steric shields to blocking penicillin from reaching the β-lactamase active site.

steric shields are the two *ortho*-methoxy groups on the aromatic ring.

Methicillin is by no means an ideal drug, however. With no electron-withdrawing group on the side chain, it is acid sensitive and has to be injected. It also shows poor activity against many other bacterial strains. Better β-lactamase-resistant agents have since been developed (see Box 19.5), and methicillin is no longer used clinically. **Nafcillin** (Fig. 19.27) is a penicillin that is resistant to β-lactamase enzymes and contains a naphthalene ring which acts as its steric shield. **Temocillin** is another β-lactamase-resistant penicillin and is interesting in that it has a 6-methoxy group present (section 19.5.1.4).

In general, β-lactamase-resistant penicillins are kept as 'reserve troops'. They are only introduced into the fray if an infection proves resistant to a broad-spectrum penicillin as a result of the presence of a β-lactamase enzyme (e.g. penicillin-resistant *S. aureus* and *Staphylococcus epidermidis*).

Unfortunately, 95% of *S. aureus* strains detected in hospitals have become resistant to methicillin and the other β-lactamase-resistant penicillins as a result of mutations to the transpeptidase enzyme. These bacteria are referred to as MRSA. The abbreviation stands for methicillin-resistant *S. aureus*, but the term applies to all the β-lactamase-resistant penicillins, not just methicillin.

Broad-spectrum penicillins

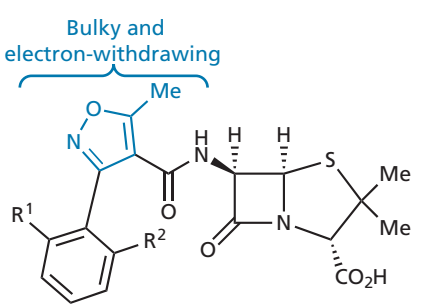
There are a variety of factors affecting whether a particular bacterial strain will be susceptible to a penicillin. The spectrum of activity shown by any penicillin depends on its structure, its ability to cross the cell membrane of Gram-negative bacteria, its susceptibility to β-lactamases, its affinity for the transpeptidase target enzyme, and the rate at which it is pumped back out of cells by Gram-negative organisms. All these factors vary in importance across different bacterial species and so there are no clear-cut tactics which can be used

BOX 19.5 The isoxazolyl penicillins

The incorporation of an isoxazolyl ring into the penicillin side chain led to orally active compounds which were stable to the β-lactamase enzyme of *S. aureus*. The isoxazolyl ring acts as the steric shield but it is also electron-withdrawing, giving the structure acid stability.

Oxacillin, **cloxacillin**, **flucloxacillin**, and **dicloxacillin** are all useful against *S. aureus* infections. The only difference between them is the type of halogen substitution on the aromatic ring. These substituents affect pharmacokinetic properties such as absorption and plasma protein binding.

Test your understanding and practise your molecular modelling with Exercise 19.3.



Oxacillin	$R^1 = R^2 = H$
Cloxacillin	$R^1 = Cl$, $R^2 = H$
Flucloxacillin	$R^1 = Cl$, $R^2 = F$
Dicloxacillin	$R^1 = Cl$, $R^2 = Cl$

Incorporation of a five-membered heterocycle into a penicillin side chain.

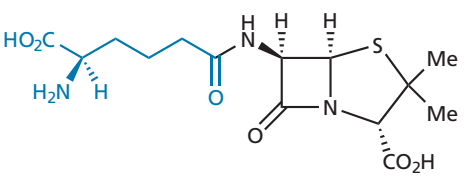
BOX 19.6 Clinical aspects of β-lactamase-resistant penicillins

Methicillin was useful in the 1960s against penicillin-resistant *S. aureus* infections. However, it is no longer used clinically. **Nafcillin** has more intrinsic activity than methicillin against staphylococci and streptococci, and is administered by injection. **Temocillin** is not active against Gram-positive bacteria, or bacteria with altered penicillin-binding proteins. It should be reserved for the treatment of infections caused by β-lactamase producing strains of Gram-negative bacteria, including those resistant to third generation cephalosporins. It is used for the treatment of septicaemia, urinary tract infections, and lower respiratory tract infections caused by susceptible Gram-negative bacteria.

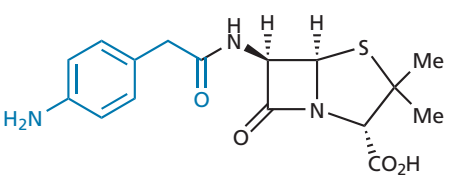
Oxacillin, **cloxacillin**, **flucloxacillin**, and **dicloxacillin** are all useful agents against *S. aureus* infection. Cloxacillin is better absorbed through the gut wall than oxacillin, whereas flucloxacillin is less bound to plasma protein, resulting in higher levels of the free drug in the blood supply. They all show inferior activity to the original penicillins if they are used against bacteria that lack the β-lactamase enzyme. They are also inactive against Gram-negative bacteria. Flucloxacillin is the drug of choice for the treatment of penicillin-resistant staphylococcal infections in the ear. **Co-fluampicil** is a combination of flucloxacillin with ampicillin, which is used against streptococcal or staphylococcal infections.

to improve the spectrum of activity. Consequently, the search for broad-spectrum antibiotics was one of trial and error which involved making a huge variety of analogues. These changes were again confined to variations in the side chain and gave the following results:

- hydrophobic groups on the side chain (e.g. penicillin G) favour activity against Gram-positive bacteria, but result in poor activity against Gram-negative bacteria;
- if the hydrophobic character is increased, there is little effect on Gram-positive activity, but activity drops even further against Gram-negative bacteria;
- hydrophilic groups on the side chain have little effect on Gram-positive activity (e.g. **penicillin T**) or cause a reduction of activity (e.g. **penicillin N**) (Fig. 19.28); however, they lead to an increase in activity against Gram-negative bacteria;
- enhancement of Gram-negative activity is found to be greatest if the hydrophilic group (e.g. NH_2, OH, CO_2H) is attached to the carbon that is α to the carbonyl group on the side chain.



Penicillin N



Penicillin T

FIGURE 19.28 Effect of side chain hydrophilic groups on antibacterial activity.

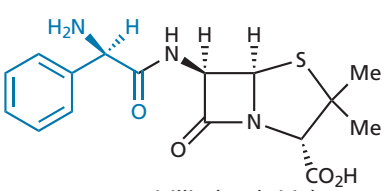
Those penicillins having useful activity against both Gram-positive and Gram-negative bacteria are known as **broad-spectrum antibiotics** (Box 19.8)**.** There are three classes of broad-spectrum antibiotics, all of which have an α–hydrophilic group which aids the passage of these penicillins through the porins of the Gram-negative bacterial outer membrane.

Broad-spectrum penicillins: the aminopenicillins

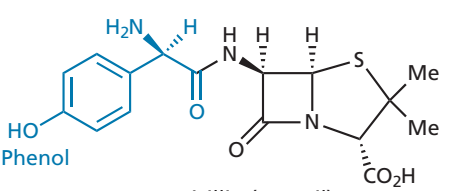
Ampicillin (Fig. 19.29; Beechams, 1964) and **amoxicillin** are orally active compounds that have a very similar structure, and are commonly used as a first line of defence against infection. Both compounds are acid resistant because of the presence of the electron-withdrawing amino group. There are no steric shields present and so these agents are sensitive to β-lactamase enzymes. Both structures are poorly absorbed through the gut wall as both the amino group and the carboxylic group are ionized. This problem can be alleviated by using a prodrug where one of the polar groups is masked with a protecting group which can be removed metabolically once the prodrug has been absorbed (Box 19.7).

Broad-spectrum penicillins: the carboxypenicillins

Carbenicillin (Fig. 19.30) was the first example of this class of compounds. It shows a broad spectrum of activity due to the hydrophilic carboxylic acid group (ionized at

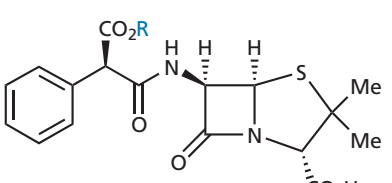


Ampicillin (Penbritin)

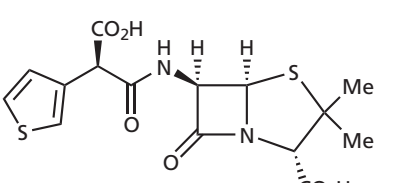


Amoxicillin (Amoxil)

FIGURE 19.29 Broad-spectrum penicillins—the aminopenicillins.



R = H Carbenicillin
R = Ph Carfecillin
R = Indanyl carbenicillin



Ticarcillin

FIGURE 19.30 Carboxypenicillins.

BOX 19.7 Ampicillin prodrugs

Pivampicillin, talampicillin, and bacampicillin are prodrugs of ampicillin (Fig. 1). In all three cases, the esters used to mask the carboxylic acid group seem rather elaborate and one may ask why a simple methyl ester is not used. The answer is that methyl esters of penicillins are not metabolized in humans. The bulky penicillin skeleton is so close to the ester that it acts as a steric shield and prevents the esterase enzymes that catalyse this reaction from accepting the penicillin ester as a substrate.

Fortunately, acyloxymethyl esters *are* susceptible to esterases. These 'extended' esters contain a second ester group further away from the penicillin nucleus, which is more exposed to attack. The hydrolysis products are inherently unstable and decompose spontaneously to release formaldehyde and reveal the free carboxylic acid (Fig. 2). The release of formaldehyde is not ideal, as it is a toxic chemical. However, it is formed naturally in the body through enzymatic demethylation of various compounds found in the diet and the levels produced from the prodrugs described cause little problem. Moreover, the drugs are only taken for a short period of time.

Such extended esters can be used to prepare prodrugs of other penicillins, but one has to be careful that one does not go to the other extreme and make the penicillin too lipophilic. For example, the 1-acyloxyalkyl ester of penicillin G is so lipophilic that it has poor solubility in water. Fortunately, the problem can be avoided easily by making the extended ester more polar (e.g. by attaching valine as in Fig. 3).

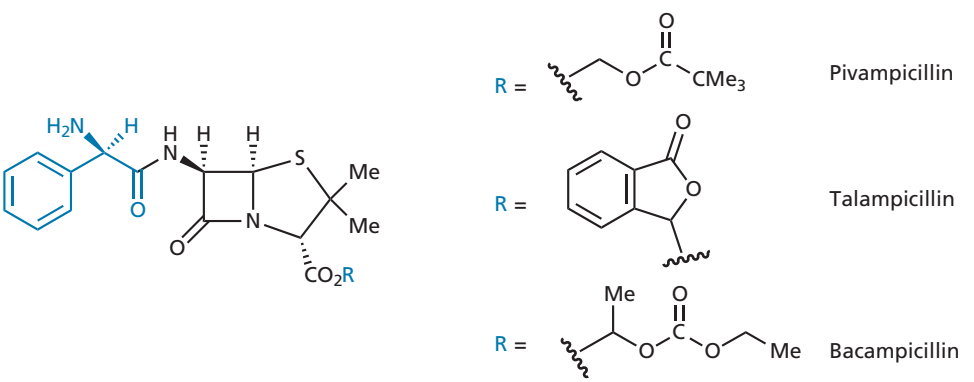


FIGURE 1 Prodrugs used to aid absorption of ampicillin through the gut wall.

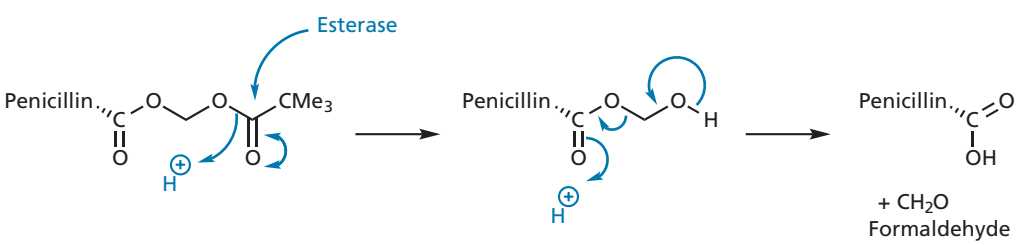


FIGURE 2 Mechanism by which acyloxymethyl esters are hydrolysed.

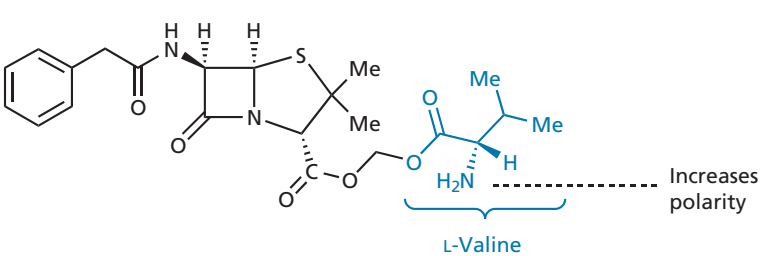


FIGURE 3 Polar extended ester for penicillin G.

pH 7) on the side chain. The stereochemistry of this group is important and only one of the two enantiomers is active.

Carfecillin and **indanyl carbenicillin** (Fig. 19.30) are prodrugs for carbenicillin and show an improved absorption through the gut wall. Aryl esters are better than alkyl esters as the former are more chemically susceptible to hydrolysis, because of the electron-withdrawing inductive effect of the aryl ring. An extended ester is not required in this case as the aryl ester is further from the β-lactam ring and is not shielded (see Box 10.7). **Ticarcillin** is similar in structure to carbenicillin, but has a thiophene ring in place of the phenyl group.

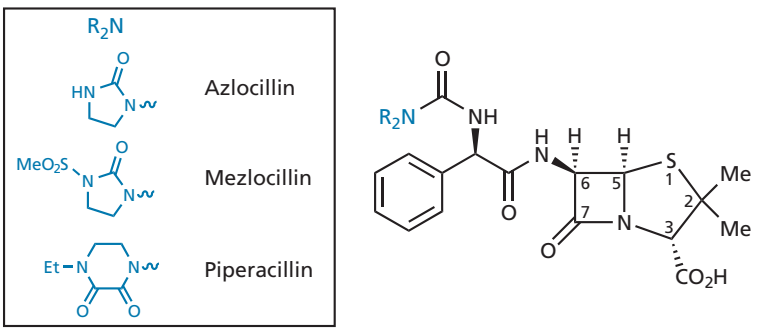


FIGURE 19.31 Ureidopenicillins.

BOX 19.8 Clinical aspects of broad-spectrum penicillins

Ampicillin and **amoxicillin** have a similar spectrum of activity to penicillin G, but are more active against Gram-negative cocci and enterobacteria. They are non-toxic and can be taken orally, but they are sensitive to β-lactamases and are inactive against *P. aeruginosa*. Some patients get diarrhoea when they take these penicillins. This is a result of poor absorption from the gut, with ampicillin being more poorly absorbed than amoxicillin. If penicillins are used at high doses for prolonged periods, they abolish the normal gut microflora and this allows the colonization of resistant Gram-negative bacilli or fungi, which cause the intestinal problems. Ampicillin is currently used to treat sinusitis, bronchitis and a variety of other infections, including oral, ear, and urinary tract infections. Amoxicillin has been used in the treatment of bronchitis, pneumonia, typhoid, gonorrhoea, Lyme disease, and urinary tract infections. Its spectrum of activity is increased when administered with **clavulanic acid** (section 19.5.4.1).

Carbenicillin was the first penicillin to show activity against *P. aeruginosa*. Compared with ampicillin, it is active against a wider range of Gram-negative bacteria and was used particularly against penicillin-resistant strains. However, it is less active than ampicillin against various other bacterial strains and requires high dose levels. Toxic side effects are observed and the drug shows a marked reduction in activity against Gram-positive bacteria. It is also acid sensitive and has to be injected. Better penicillins, such as the ureidopenicillins, have since been developed and so the use of carbenicillin is now discouraged.

Carfecillin and **indanyl carbenicillin** proved useful for the treatment of urinary tract infections, but have generally been superseded by fluoroquinolone antibacterial agents (section 19.8.1).

Ticarcillin is administered by injection and has an identical antibacterial spectrum to carbenicillin. However, it has the advantage that smaller doses can be used. It is also 2–4 times more effective against *P. aeruginosa* and has fewer side effects. The drug is used mainly against infections due to *Pseudomonas* and *Proteus* species, and is currently administered with clavulanic acid to broaden its spectrum of activity (section 19.5.4.1).

Ureidopenicillins are generally more active than the carboxypenicillins against streptococci and *Haemophilus* species. They show similar activity against Gram-negative aerobic rods such as *P. aeruginosa,* but are generally more active against other Gram-negative bacteria. Unfortunately, they have to be injected. Examples include **azlocillin**, which is 8–16 times more active than carbenicillin against *P. aeruginosa* and is used primarily for the treatment of infections caused by that organism. It is susceptible to β-lactamases. **Mezlocillin** has a similar spectrum of activity to carbenicillin, but is more active because it has a higher affinity for transpeptidases and can cross the outer membrane of Gram-negative bacteria more effectively. **Piperacillin** is similar to ampicillin in its activity against Gram-positive species. It also has good activity against anaerobic species of both cocci and bacilli, and can be used against a variety of infections. It is more active than ticarcillin against *P. aeruginosa*. Piperacillin can be administered alongside **tazobactam** to widen its spectrum of activity (section 19.5.4.2).

Broad-spectrum penicillins: the ureidopenicillins

Ureidopenicillins (Fig. 19.31) are the newest class of broad-spectrum penicillins and have a urea functional group at the α-position. Generally, they have better properties than the **carboxypenicillins** and have largely replaced them in the clinic.

19.5.1.9 Synergism of penicillins with other drugs

There are several examples in medicinal chemistry where the presence of one drug enhances the activity of another. In many cases this can be dangerous, leading to an effective overdose of the enhanced drug. In some cases, though, it can be useful. There are two interesting examples where the activity of penicillin has been enhanced by the presence of another drug.

One of these is the effect of clavulanic acid, described in Section 19.5.4.1. The other is the administration of penicillins with a compound called **probenecid** (Fig. 19.32). Probenecid is a moderately lipophilic carboxylic acid that can block facilitated transport of penicillin through the kidney tubules. In other words, probenecid slows down the rate at which penicillin is excreted, by competing with it in the excretion mechanism. Probenecid also competes with penicillin for binding sites on albumin. As a result, penicillin levels in the bloodstream are enhanced and the antibacterial activity increases—a useful tactic if faced with a particularly resistant bacterium.

KEY POINTS

- Penicillins have a bicyclic structure consisting of a β-lactam ring fused to a thiazolidine ring. The strained β-lactam ring reacts irreversibly with the transpeptidase enzyme responsible for the final cross-linking of the bacterial cell wall.
- Penicillin analogues can be prepared by fermentation or by a semi-synthetic synthesis from 6-aminopenicillanic acid. Variation of the penicillin structure is limited to the acyl side chain.
- Penicillins can be made more resistant to acid conditions by incorporating an electron-withdrawing group into the acyl side chain.
- Steric shields can be added to penicillins to protect them from bacterial β-lactamase enzymes.
- Broad spectrum activity is associated with the presence of an α-hydrophilic group on the acyl side chain of penicillin.
- Prodrugs of penicillins are useful in masking polar groups and improving absorption from the gastrointestinal tract. Extended esters are used which undergo enzyme-catalysed hydrolysis to produce a product which degrades spontaneously to release the penicillin.
- Probenecid can be administered with penicillins to hinder the excretion of penicillins.

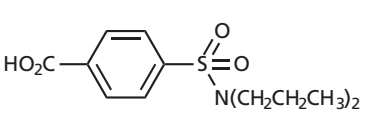


FIGURE 19.32 Probenecid.

19.5.2 Cephalosporins

19.5.2.1 Cephalosporin C

Discovery and structure of cephalosporin C

The second major group of β-lactam antibiotics to be discovered were the cephalosporins. The first cephalosporin (**cephalosporin C**) was derived from a fungus obtained in the mid 1940s from sewer waters on the island of Sardinia. This was the work of an Italian professor who noted that the waters surrounding the sewage outlet periodically cleared of microorganisms. He reasoned that an organism might be producing an antibacterial substance and so he collected samples and managed to isolate a fungus called *Cephalosporium acremonium* (now called *Acremonium chrysogenum*). The crude extract from this organism was shown to have antibacterial properties and, in 1948, workers at Oxford University isolated cephalosporin C, but it was not until 1961 that the structure was established by X-ray crystallography.

The structure of cephalosporin C (Fig. 19.33) has similarities to that of penicillin in that it has a bicyclic system containing a four-membered β-lactam ring, but this time the β-lactam ring is fused to a six-membered dihydrothiazine ring. Nevertheless, cephalosporins are derived from the same biosynthetic precursors as penicillin (i.e. cysteine and valine) (Fig. 19.34).

Properties of cephalosporin C

Cephalosporin C is not particularly potent compared with penicillins (1/1000 the activity of penicillin G), but the antibacterial activity it *does* have is more evenly directed against Gram-negative and Gram-positive bacteria. Another in-built advantage of cephalosporin C is

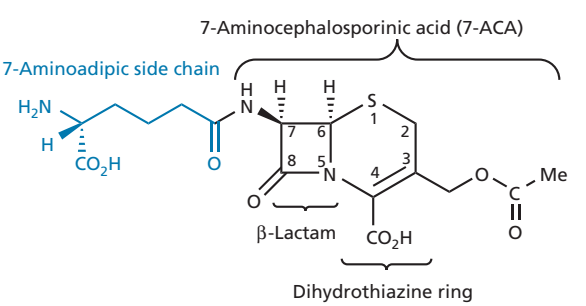


FIGURE 19.33 Cephalosporin C.

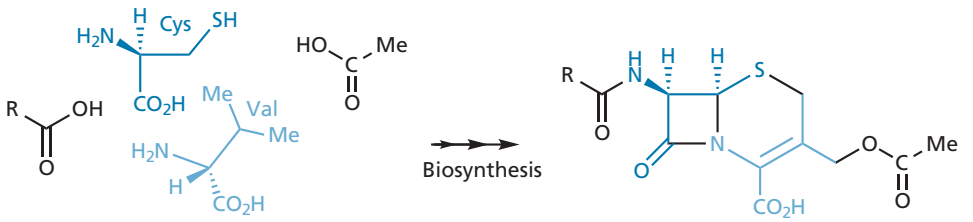


FIGURE 19.34 Biosynthetic precursors of cephalosporin C.

its greater resistance to acid hydrolysis and β-lactamase enzymes. It is also less likely to cause allergic reactions. Therefore, cephalosporin C was seen as a useful lead compound for the development of further broad-spectrum antibiotics, hopefully with increased potency.

Structure–activity relationships of cephalosporin C

Many analogues of cephalosporin C have been made which demonstrate the importance of the β-lactam ring within the bicyclic system, an ionized carboxylate group at position 4, and the acylamino side chain at position 7. These results tally closely with those obtained for the penicillins. The strain effect of a 6-membered ring fused to a 4-membered ring is less than for penicillin, but this is partially offset by the effect of the acetyloxy group at position 3. This can act as a good leaving group in the inhibition mechanism (Fig. 19.35).

There is a limited number of places where modifications can be made (Fig. 19.36), but there are more possibilities than with penicillins. These are as follows;

- variations of the 7-acylamino side chain;
- variations of the 3-acetoxymethyl side chain;
- extra substitution at carbon 7.

19.5.2.2 Synthesis of cephalosporin analogues at position 7

Access to analogues with varied side chains at position 7 initially posed a problem. Unlike penicillins, it proved impossible to obtain cephalosporin analogues by fermentation. Similarly, it was not possible to obtain **7-ACA (7-aminocephalosporinic acid)** either by fermentation or by enzymatic hydrolysis of cephalosporin C, thus preventing the semi-synthetic approach analogous to the preparation of penicillins from 6-APA (section 19.5.1.6).

Therefore, a way had to be found of obtaining 7-ACA from cephalosporin C by chemical hydrolysis. This is no easy task, as a secondary amide has to be hydrolysed in the presence of a highly reactive β-lactam ring. Normal hydrolytic procedures are not suitable and so a special method had to be worked out (Fig. 19.37).

The first step of the procedure requires the formation of an imino chloride by the mechanism shown in Fig. 19.38. This is only possible for the secondary amide group, as ring constraints prevent the β-lactam nitrogen forming a double bond within the β-lactam ring. The imino chloride can then be treated with an alcohol to give an imino ether. This functional group is more susceptible to hydrolysis than the β-lactam ring, and so treatment with aqueous acid successfully gives the desired 7-ACA which can then be acylated to give a range of analogues.

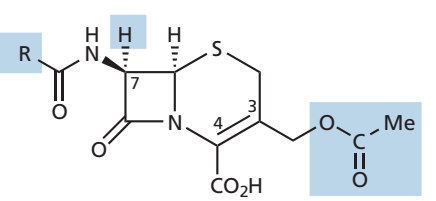


FIGURE 19.36 Positions for possible modification of cephalosporin C. The shading indicates positions which can be varied.

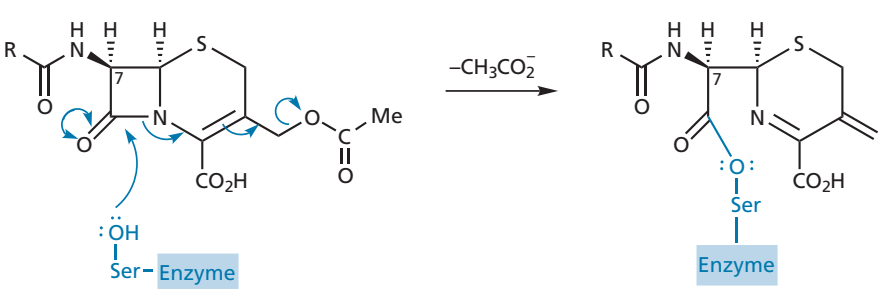


FIGURE 19.35 Mechanism by which cephalosporins inhibit the transpeptidase enzyme.

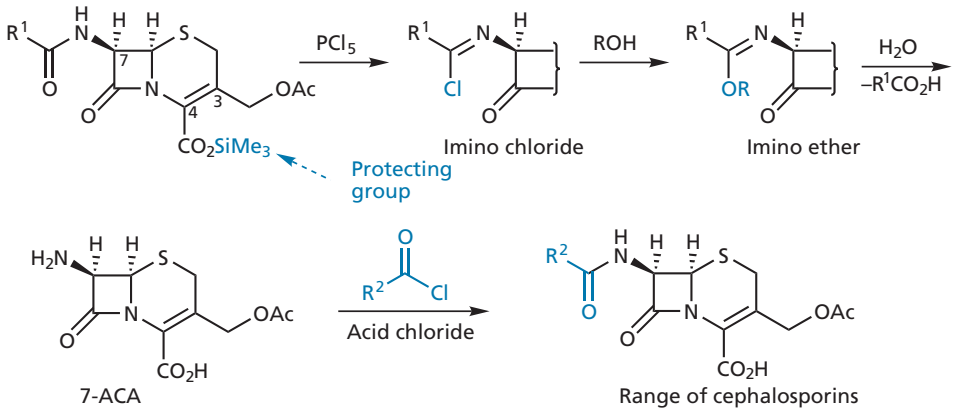


FIGURE 19.37 Synthesis of 7-ACA and cephalosporin analogues.

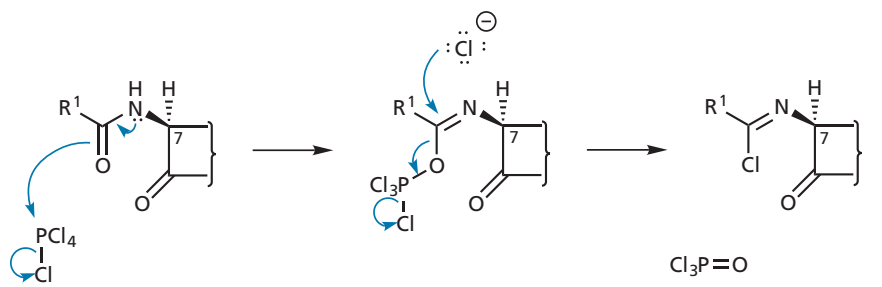


FIGURE 19.38 Mechanism for imino chloride formation.

19.5.2.3 First-generation cephalosporins

Examples of first-generation cephalosporins include **cephalothin**, **cephaloridine**, **cefalexin**, and **cefazolin** (Figs 19.39–19.42). In general, they have a lower activity than comparable penicillins, but a better range. Most are poorly absorbed through the gut wall and have to be injected. As with penicillins, the appearance of resistant organisms has posed a problem, particularly with Gram-negative organisms. These contain β-lactamases which are more effective than the β-lactamases of Gram-positive organisms. Steric shields are successful in protecting cephalosporins from these β-lactamases, but also prevent them from inhibiting the transpeptidase target enzymes.

One of the most commonly used first-generation cephalosporins was **cephalothin** (Fig. 19.39). A disadvantage with cephalothin is the fact that the acetyloxy group at position 3 is readily hydrolysed by esterase enzymes to give the less active alcohol (Fig. 19.40). The acetyloxy group is important to the mechanism of inhibition and acts as a good leaving group, whereas the alcohol is a much poorer leaving group. Therefore, it would be useful if this metabolism could be blocked to prolong activity. Replacing the ester with a metabolically stable pyridinium group gives **cephaloridine** (Fig. 19.41). The pyridine can still act as a good leaving group for the inhibition mechanism, but is not cleaved by esterases. Cephaloridine exists as a zwitterion and is soluble in water, but, like most first-generation cephalosporins, it is poorly absorbed through the gut wall and has to be injected.

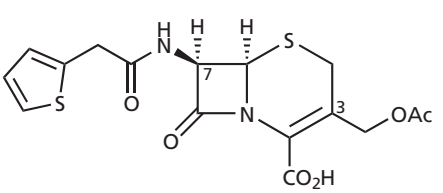


FIGURE 19.39 Cephalothin.

Cefalexin (Fig. 19.41) has a methyl substituent at position 3 (Box 19.9) which appears to help oral absorption. A methyl group would normally be bad for activity as it is not a good leaving group. However, the presence of a hydrophilic amino group at the α-carbon of the 7-acylamino side chain in cefalexin helps to restore activity and cephalexin is one of the few cephalosporins which is orally active. The mechanism of absorption through the gut wall is poorly understood and it is not clear why the 3-methyl group is so advantageous for absorption. **Cefazolin** (Fig. 19.42) is another example of a first-generation cephalosporin.

FIGURE 19.40 Metabolic hydrolysis of cephalothin.

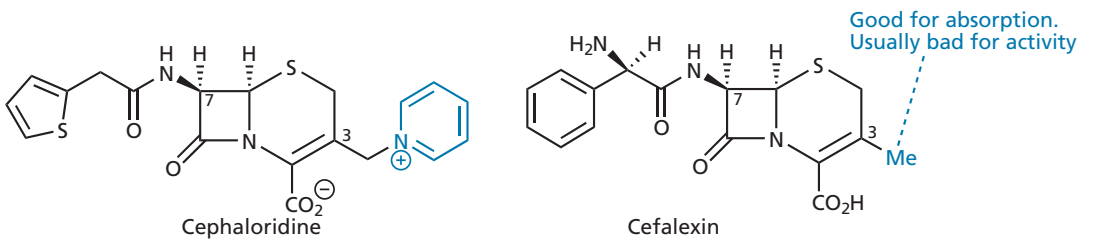
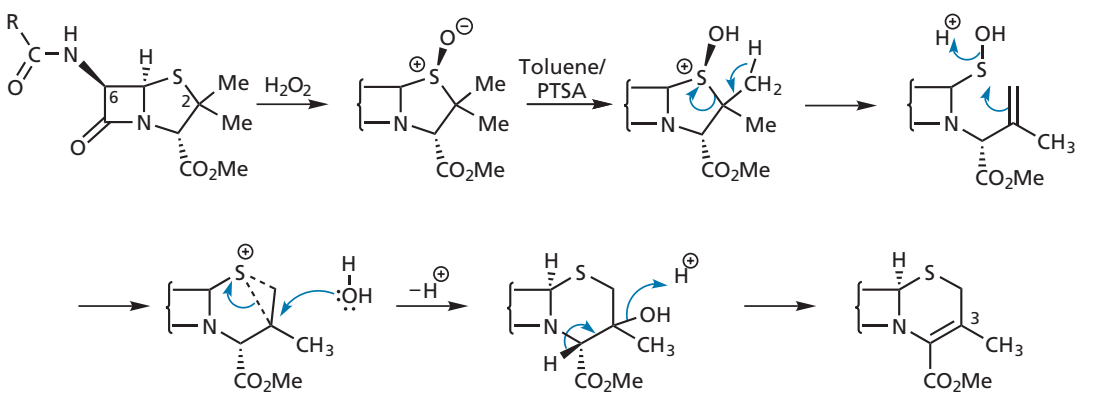


FIGURE 19.41 Cephaloridine and cefalexin.

BOX 19.9 Synthesis of 3-methylated cephalosporins

The synthesis of 3-methylated cephalosporins involves the use of a penicillin starting material as shown below. The synthesis, which was first demonstrated by Eli Lilly Pharmaceuticals, involves oxidation of sulphur followed by an acid-catalysed ring expansion, where the five-membered thiazolidine ring in penicillin is converted to the six-membered dihydrothiazine ring in cephalosporin.



Synthesis of 3-methylated cephalosporins from a penicillin.

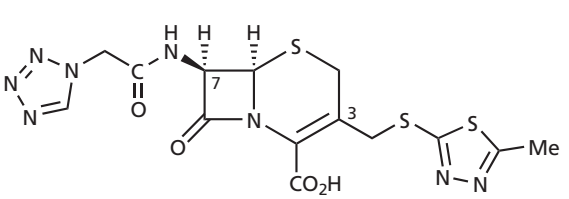


FIGURE 19.42 Cefazolin.

19.5.2.4 Second-generation cephalosporins

Cephamycins

Cephamycins contain a methoxy substituent at position 7, which has proved advantageous. The parent compound **cephamycin C** (Fig. 19.43) was isolated from a culture of *Streptomyces clavuligerus* and was the first β-lactam to be isolated from a bacterial source. Modification of the side

chain gave **cefoxitin** (Fig. 19.43), which showed a broader spectrum of activity than most first-generation cephalosporins. This is due to greater resistance to β-lactamase enzymes, which may be due to the steric hindrance provided by the methoxy group. Cefoxitin shows good metabolic stability to esterases owing to the presence of the urethane group at position 3, rather than an ester (section 14.2.2).

Test your understanding and practise your molecular modelling with Exercise 19.4.

Oximinocephalosporins

The development of **oximinocephalosporins** has been a major advance in cephalosporin research. These structures contain an iminomethoxy group at the α-position of the acyl side chain, which significantly increases the stability of cephalosporins against the β-lactamases produced by some organisms (e.g. *Haemophilus influenza*). The first useful agent in this class of compounds was **cefuroxime** (Fig. 19.44), which, like cefoxitin, has an increased resistance to β-lactamases and mammalian esterases. Unlike cefoxitin, cefuroxime retains activity against streptococci and, to a lesser extent, staphylococci.

19.5.2.5 Third-generation cephalosporins

Replacing the furan ring of the aforesaid oximinocephalosporins with an aminothiazole ring enhances the penetration of cephalosporins through the outer membrane of Gram-negative bacteria, and may also increase affinity for the transpeptidase enzyme. As a result, third-generation cephalosporins containing this ring have a marked increase in activity against these bacteria. A variety of such structures have been prepared, such as **ceftazidime**, **cefotaxime**, **ceftizoxime**, and **ceftriaxone** (Figs 19.44 and 19.45), with different substituents at position 3 to vary the pharmacokinetic properties. They play a major role in antimicrobial therapy because of their activity against Gram-negative bacteria, many of which are resistant to other β-lactams. As such infections are uncommon outside hospitals, physicians are discouraged from prescribing these drugs routinely and they are viewed as 'reserve troops' to be used for troublesome infections which do not respond to the more commonly prescribed β-lactams.

19.5.2.6 Fourth-generation cephalosporins

Cefepime and **cefpirome** (Fig. 19.45) are oximinocephalosporins which have been classed as fourth-generation cephalosporins. They are zwitterionic compounds having a positively charged substituent at position 3 and a negatively charged carboxylate group at position 4. This property appears to radically enhance the ability of these compounds to penetrate the outer membrane of Gram-negative bacteria. They are also found to have a good affinity for the transpeptidase enzyme and a low affinity for a variety of β-lactamases.

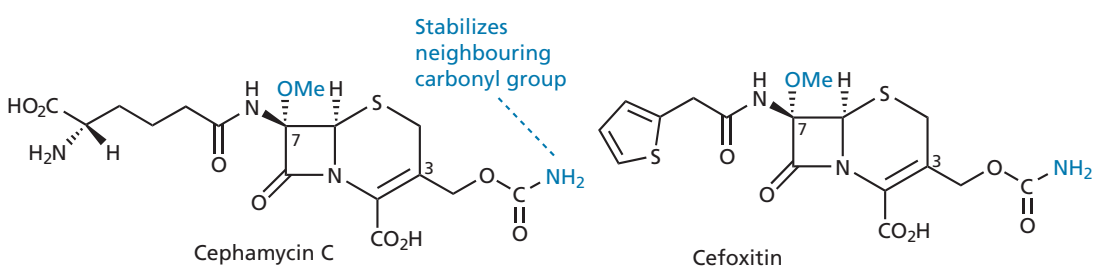


FIGURE 19.43 Cephamycin C and cefoxitin.

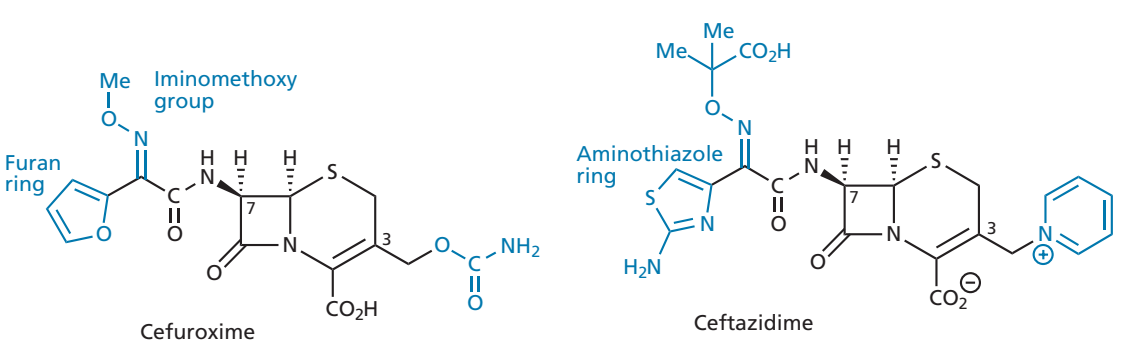


FIGURE 19.44 Oximinocephalosporins.

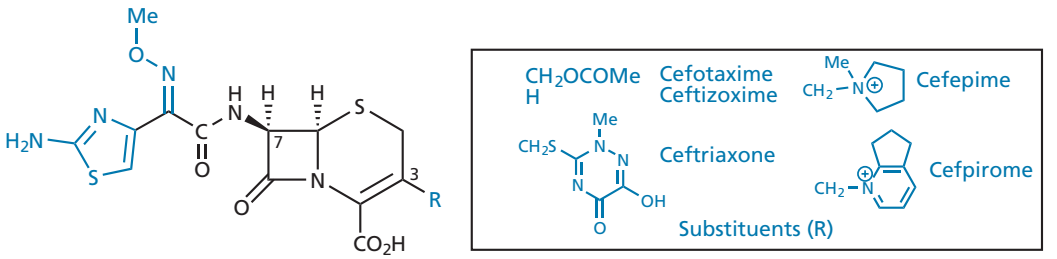


FIGURE 19.45 Third- and fourth-generation oximinocephalosporins.

19.5.2.7 Fifth-generation cephalosporins

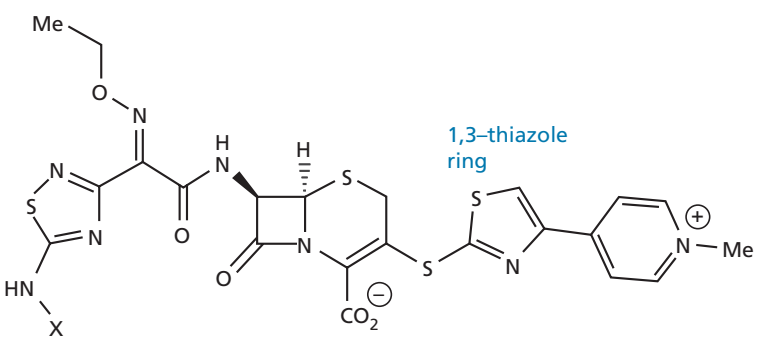
Ceftaroline fosamil (Fig. 19.46) is a fifth-generation cephalosporin that has activity against various strains of MRSA and multi-resistant *Streptococcus pneumonia* (MDRSP). It acts as a prodrug for **ceftaroline**, and the 1,3-thiazole ring is thought to be important for its activity against MRSA.

19.5.2.8 Resistance to cephalosporins

The activity of a specific cephalosporin against a particular bacterial cell is dependent on the same factors as those for penicillins. i.e. the ability to reach the transpeptidase enzyme, stability to any β-lactamases which might be present, and the affinity of the antibiotic for the target. For example, most cephalosporins (with the exception of cephaloridine) are stable to the β-lactamase produced by *S. aureus* and can reach the transpeptidase enzyme without difficulty. Therefore, the relative ability of cephalosporins to inhibit *S. aureus* comes down to their affinity for the target transpeptidase enzyme. Agents such as the cephamycins and ceftazidime have poor affinity, whereas other cephalosporins have a higher affinity. The MRSA organism contains a modified transpeptidase enzyme (**PBP2a**) for which both penicillins and cephalosporins have poor affinity.

KEY POINTS

- Cephalosporins contain a strained β-lactam ring fused to a dihydrothiazine ring.
- In general, first-generation cephalosporins offer advantages over penicillins in that they have greater stability to acid conditions and β-lactamases, and have a good ratio of activity against Gram-positive and Gram-negative bacteria. However, they have poor oral availability and are generally lower in activity.
- Variation of the 7-acylamino side chain alters antimicrobial activity, whereas variation of the side chain at position 3 predominantly alters the metabolic and pharmacokinetic properties of the compound. Introduction of a methoxy substitution at C-7 is possible.
- Semisynthetic cephalosporins can be prepared from 7-aminocephalosporanic acid (7-ACA).
- 7-ACA is obtained from the chemical hydrolysis of cephalosporins. This requires prior activation of the side chain to make it more reactive than the β-lactam ring.
- Deacetylation of cephalosporins occurs metabolically to produce inactive metabolites. Metabolism can be blocked by replacing the susceptible acetoxy group with metabolically stable groups.



Ceftaroline; X = H
Ceftaroline fosamil; X = $P(=O)(OH)_2$

FIGURE 19.46 Ceftaroline and ceftaroline fosamil.

BOX 19.10 Clinical aspects of cephalosporins

In general, cephalosporins are useful broad-spectrum antibacterial agents for the treatment of septicaemia, pneumonia, meningitis, biliary tract infections, peritonitis, and urinary tract infections. **Cephalosporin C** itself has been used in the treatment of urinary tract infections, as it is found to concentrate in the urine and survive the body's hydrolytic enzymes.

First-generation cephalosporins

First-generation cephalosporins have good activity against Gram-positive cocci and they can be used to treat some community-derived Gram-negative infections (i.e. infections not caught in a hospital). They can also be used against *S. aureus* and streptococcal infections when penicillins have to be avoided. **Cephalothin** is more active than penicillin G against some Gram-negative bacteria and is less likely to cause allergic reactions. It can also be used against β-lactamase producing *S. aureus* strains.

Cefalexin is useful for the treatment of urinary tract infections which do not respond to other drugs or which occur in pregnancy. It is also useful in treating infections of the respiratory tract, ear, skin, and mouth. **Cefazolin** is recommended for use as a prophylactic to prevent infection when surgical procedures are used to implant foreign bodies.

Second-generation cephalosporins

In general, the second-generation cephalosporins have variable activity against Gram-positive cocci, but increased activity against Gram-negative bacteria. **Cefoxitin** is active against bowel flora, including *Bacteroides fragilis*, and was once recommended for peritonitis. **Cefuroxime** has a wide spectrum of activity and is useful against organisms which have become resistant to penicillin. However, it is not active against 'difficult' bacteria, such as *P. aeruginosa*. It is used clinically against *Neisseria gonorrhoeae* and respiratory infections caused by *H. influenza*, *Moraxella catarrhalis*, and susceptible strains of *S. pneumoniae*. It is also used for surgical prophylaxis, as well as for the treatment of Lyme disease. **Cefotaxime** is used in surgical prophylaxis and for the treatment of gonorrhoea, meningitis, and infections caused by *Haemophilus epiglottis*.

Third-generation cephalosporins

Third-generation cephalosporins have good activity against Gram-negative bacteria, but vary in their activity against Gram-positive cocci. The ability to attack *P. aeruginosa* also varies from structure to structure, and they lack activity against the MRSA organisms and *Enterobacter* species. **Ceftazidime** is an injectable cephalosporin which has excellent activity against *P. aeruginosa*, as well as other Gram-negative bacteria. Because the drug can cross the blood–brain barrier it can be used to treat meningitis. Compared with the other aminothiazole structures, ceftazidime has good activity against streptococci, but loses activity against strains of methicillin-susceptible *S. aureus*. This is because of a decreased binding affinity for the transpeptidase enzyme present in *S. aureus*. **Ceftriaxone** is used for surgical prophylaxis and as a prophylactic for meningococcal meningitis.

Fourth and fifth-generation cephalosporins

Fourth-generation cephalosporins have activity against Gram-positive cocci and a broad array of Gram-negative bacteria, including *P. aeruginosa* and many of the enterobacterial species. **Cefpirome** is administered as an intravenous injection or infusion, and has been used against a variety of sensitive Gram-positive and Gram-negative bacteria. **Ceftaroline fosamil** has been licensed for the treatment of bacterial pneumonia and acute bacterial skin infections.

- A methyl substituent at position 3 is good for oral absorption but bad for activity unless a hydrophilic group is present at the α-position of the acyl side chain.
- 3-Methylated cephalosporins can be synthesized from penicillins.
- Cephamycins are cephalosporins containing a methoxy group at position 7.
- Oximinocephalosporins have resulted in several generations of cephalosporins with increased potency and a broader spectrum of activity, particularly against Gram-negative bacteria.

19.5.3 Other β-lactam antibiotics

Although penicillins and cephalosporins are the best known and most researched β-lactams, there are other β-lactam structures which are of great interest in the antibacterial field.

19.5.3.1 Carbapenems

Thienamycin (Fig. 19.47) was the first example of this class of compounds and was isolated from *Streptomyces cattleya* in 1976. It is potent, with an extraordinarily

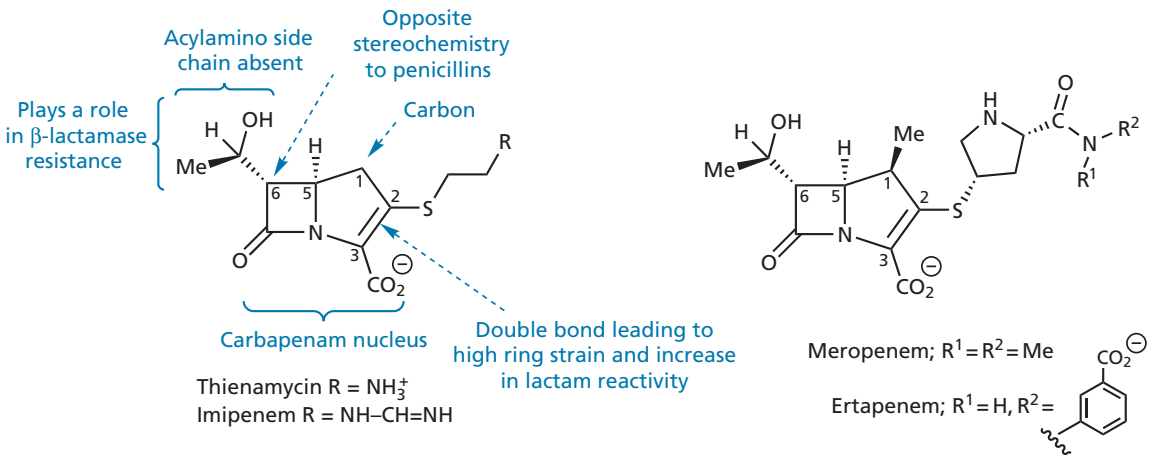


FIGURE 19.47 Carbapenems.

broad range of activity against Gram-positive and Gram-negative bacteria, including *P. aeruginosa*. It has low toxicity and shows a high resistance to β-lactamases. This resistance has been ascribed to the presence of the hydroxyethyl side chain. Unfortunately, it shows poor metabolic and chemical stability, and is not absorbed from the gastrointestinal tract. The surprising features in thienamycin are the missing sulphur atom and acylamino side chain, both of which were thought to be essential to antibacterial activity. Furthermore, the stereochemistry of the side chain at substituent 6 is opposite from the usual stereochemistry in penicillins—another factor contributing to the resistance of this agent to β-lactamases. **Imipenem** and **meropenem** are clinically useful analogues of thienamycin (Box 19.11). Imipenem is susceptible to metabolism by a dehydropeptidase enzyme, whereas meropenem is more resistant as a result of the different substituent at position 2. **Ertapenem** was approved in 2002 and is similar in structure to meropenem. It has an extra substituent on the carbapenem ring (R^1 = Me) which provides further stability against dehydropeptidases, while the ionized benzoic acid contributes to high protein binding and prolongs the half-life of the drug such that once-daily dosing is feasible. In general, the **carbapenems** have the broadest spectrum of activity of all the β-lactam antibiotics.

BOX 19.11 Clinical aspects of miscellaneous β-lactam antibiotics

Imipenem is active against a variety of aerobic, anaerobic, Gram-positive, and Gram-negative infections, and can be effective against some infections which do not respond to cephalosporins, or infections which have become resistant to more conventional β-lactams. It can be used against hospital-acquired septicaemia and for surgical prophylaxis. The structure is metabolized by a dehydropeptidase enzyme to produce metabolites that are toxic to the kidney, but this can be alleviated by administrating the drug alongside **cilastatin**—a dehydropeptidase inhibitor which protects imipenem from metabolism. Administration is by intramuscular injection or by intravenous infusion. **Meropenem** is also effective against a variety of aerobic, anaerobic, Gram-positive, and Gram-negative infections, and is administered by intravenous injection or infusion. Meropenem is slightly less active than imipenem against Gram-positive bacteria, but is more active against Gram-negative bacteria. Unlike imipenem, meropenem is active against *P. aeruginosa* and can be administered on its own because it is more resistant to dehydropeptidases. Both meropenem and imipenem penetrate the outer membrane of Gram-negative bacteria through porins, but meropenem enters more efficiently, resulting in higher activity against these bacteria. The drug has been used to treat pneumonia, meningitis, abdominal infections, and urinary tract infections. **Ertapenem** is administered by intravenous infusion and is used for the treatment of abdominal infections, acute gynaecological infections, community-acquired pneumonia, and diabetic foot infections of the skin and soft tissue. It is also used as a prophylactic for colorectal surgery. **Aztreonam** is used against Gram-negative infections, including *P. aeruginosa*, *H. influenzae*, and *Neisseria meningitidis*. It is administered by intravenous injection and can be used safely in patients with allergies to penicillin or cephalosporins. It has no activity against Gram-positive organisms or **anaerobic bacteria** because it does not bind to the transpeptidases produced by these organisms. However, it can bind to, and inhibit, the transpeptidases produced by Gram-negative aerobic organisms.

19.5.3.2 Monobactams

Monocyclic β-lactams such as the **nocardicins** (Fig. 19.48) have been isolated from natural sources. At least seven nocardicins have been isolated by the Japanese company Fujisawa. They show moderate activity *in vitro* against a narrow group of Gram-negative bacteria, including *P. aeruginosa*. Surprisingly, they contain a single β-lactam ring, demonstrating that a fused second ring is not always essential for antibacterial activity. One explanation for this is that nocardicins might have a different mechanism of action from penicillins and cephalosporins—possibly by inhibiting a different enzyme involved in cell wall synthesis. This would help to explain why nocardicins are inactive against Gram-positive bacteria and generally show a different spectrum of activity from the other β-lactam antibiotics. They also show low levels of toxicity. **Aztreonam** (Fig. 19.48) is an example of a monobactam which has reached the clinic and was developed from a naturally occurring monobactam isolated from *Chromobacterium violaceum*.

19.5.4 β-Lactamase inhibitors

19.5.4.1 Clavulanic acid

Clavulanic acid (Fig. 19.49) was isolated from *S. clavuligerus* by Beechams in 1976. It has weak and unimportant antibiotic activity, but it is a powerful and irreversible inhibitor of most β-lactamases, which means that it is used as a sentry drug (section 14.7.1) in combination with traditional penicillins, such as amoxicillin (**Augmentin**). This allows the dose levels of amoxicillin to be decreased and also increases the spectrum of activity. However, it should be noted that there are various types of β-lactamases. Although clavulanic acid is effective against most of these, it is not effective against all. Clavulanic acid is also administered intravenously with ticarcillin as **Timentin**.

The structure of clavulanic acid was the first example of a naturally occurring compound where the β-lactam ring was not fused to a sulphur-containing ring; instead, it is fused to an oxazolidine ring. The structure is also unusual in that it does not have an acylamino side chain.

Many analogues have now been made and the essential requirements for β-lactamase inhibition are:

- a strained β-lactam ring;
- the enol ether;
- the *Z* configuration for the double bond of the enol ether (activity is reduced but not eliminated if the double bond is *E*);
- no substitution at C-6;
- (*R*)-stereochemistry at positions 2 and 5;
- the carboxylic acid group.

It is also thought that the 9-hydroxyl group is involved in a hydrogen bonding interaction with the active site of the β-lactamase enzyme. Clavulanic acid is a mechanism-based irreversible inhibitor and can be classed as a

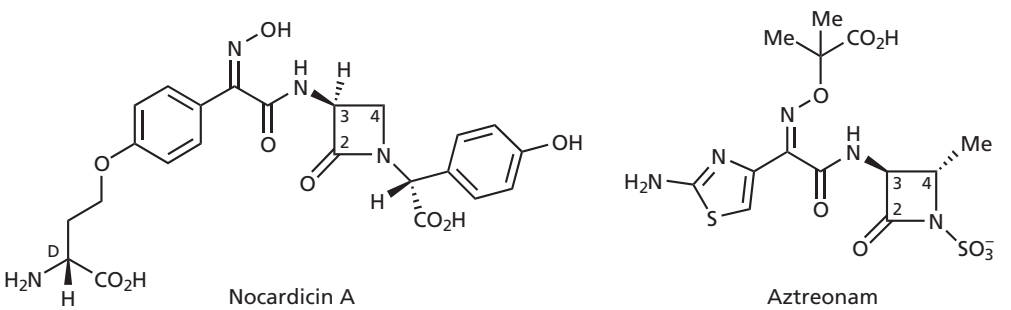


FIGURE 19.48 Monobactams.

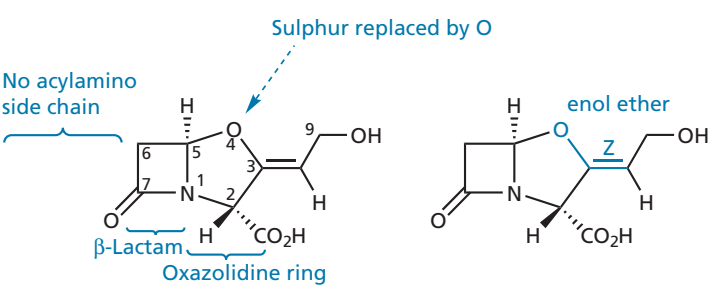


FIGURE 19.49 Clavulanic acid.

suicide substrate. The mechanism of inhibition is shown in section 7.5.

19.5.4.2 Penicillanic acid sulphone derivatives

The agents **sulbactam** and **tazobactam** have also been developed as β-**lactamase inhibitors** and are used clinically (Fig. 19.50). They, too, act as suicide substrates for β-lactamase enzymes and have similar properties. Sulbactam has a broader spectrum of activity against β-lactamases than clavulanic acid, but is less potent. It is combined with ampicillin for intravenous administration in a preparation called **Unasyn**. Tazobactam is similar to sulbactam and has a similar spectrum of activity against β-lactamases. However, its potency is more like clavulanic acid. It is administered intravenously with piperacillin in a preparation called **Tazocin** or **Zosyn**, which has the broadest spectrum of activity of the various combinations described so far.

19.5.4.3 Olivanic acids

The **olivanic acids** (e.g. **MM 13902**) (Fig. 19.51) were isolated from strains of *Streptomyces olivaceus* and are carbapenem structures like thienamycin. They are very strong inhibitors of β-lactamase—in some cases 1000 times more potent than clavulanic acid. They are also effective against the β-lactamases which break down cephalosporins and are unaffected by clavulanic acid. Unfortunately, olivanic acids lack chemical stability.

19.5.5 Other drugs which act on bacterial cell wall biosynthesis

β-Lactams are not the only antibacterial agents that inhibit cell wall biosynthesis. The antibacterial agents **vancomycin**, **D-cycloserine**, and **bacitracin** also inhibit biosynthesis, though at different stages. In order to synthesize the cell wall, *N*-acetylmuramic acid (NAM) is linked to three amino acids, then to the dipeptide D-Ala-D-Ala (Fig. 19.52). The D-Ala-D-Ala dipeptide is derived from two L-alanine units which are first racemized then linked together.

NAM, with its pentapeptide chain, is then linked to a **C55 carrier lipid** with the aid of a **translocase** enzyme and carried to the outer surface of the cell membrane, where the lipid carrier acts as an anchor to hold the glycopeptide in place for the subsequent steps. These steps involve the addition of *N*-acetylglucosamine (NAG) and a pentaglycine chain to give the complete 'building block'. A **transglycosidase** enzyme catalyses the attachment of the disaccharide building block to the growing cell wall and, at the same time, the carrier lipid is released to pick up another molecule of NAM/pentapeptide. Cross-linking between the various chains of the cell wall finally takes place, catalysed by the transpeptidase enzyme as described previously (section 19.5.1.4).

19.5.5.1 D-Cycloserine and bacitracin

D-Cycloserine (Fig. 19.53) is a simple molecule produced by *Streptomyces garyphalus*, which has broad-spectrum activity and acts within the cytoplasm to prevent the formation of D-Ala-D-Ala. It does this by mimicking the structure of D-alanine and inhibiting the enzymes **L-alanine racemase** (responsible for racemizing L-Ala to D-Ala) and **D-Ala-D-Ala ligase** (responsible for linking the two D-alanine units together).

Bacitracin is a polypeptide complex produced by *Bacillus subtilis*, which binds to the lipid carrier responsible for transporting the NAM/pentapeptide unit across the cell membrane, thus preventing it from carrying out that role.

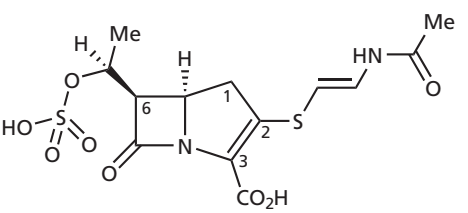
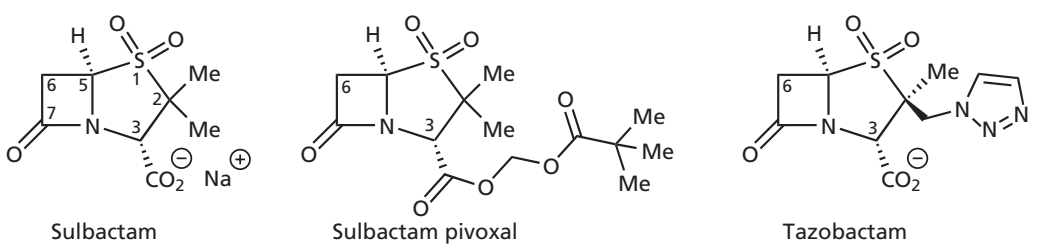


FIGURE 19.51 MM 13902.



Sulbactam | Sulbactam pivoxal | Tazobactam

FIGURE 19.50 Penicillanic acid sulphones.

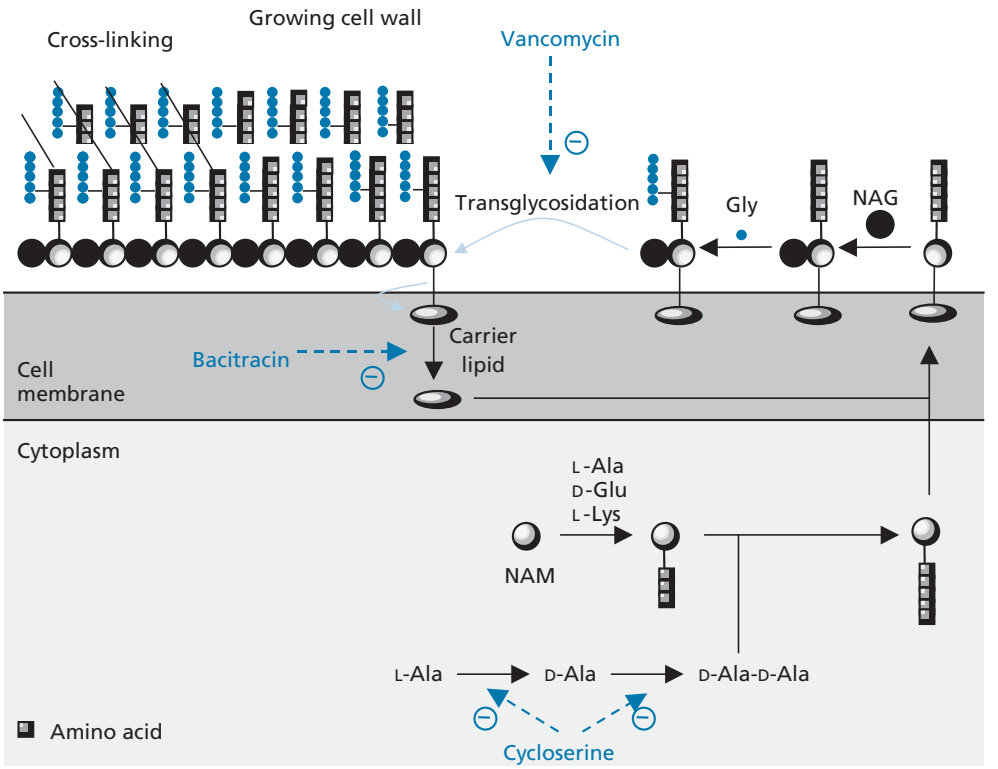
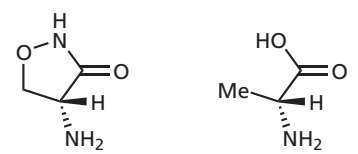


FIGURE 19.52 Cell wall biosynthesis.



D-Cycloserine D-Alanine

FIGURE 19.53 D-Cycloserine as a mimic for D-alanine.

19.5.5.2 The glycopeptides: vancomycin and vancomycin analogues

Vancomycin (Fig. 19.54) is a narrow-spectrum bactericidal glycopeptide produced by a microorganism called *Streptomyces orientalis* found in Borneo and India. Aptly, its name is derived from the verb 'to vanquish'. Vancomycin was introduced in 1956 for the treatment of infections caused by penicillin-resistant *S. aureus*, but was discontinued when methicillin became available. It has since been reintroduced and is now the main stand-by drug for treating MRSA. Vancomycin and related glycopeptides are often the last resort in treating patients with drug-resistant infections. As such, they have become extremely important and a great deal of research is currently being carried out in this area.

Vancomycin is derived biosynthetically from a linear heptapeptide containing five aromatic residues. These undergo oxidative coupling with each other to produce three cyclic moieties within the structure. Chlorination, hydroxylation, and the final addition of two sugar units then complete the structure (Fig. 19.55).

The cyclizations described transform a highly flexible heptapeptide molecule into a rigid structure that holds the peptide backbone in a fixed conformation. Moreover, there is an extra element of rigidity to the structure, which may not be apparent at first sight. The aromatic rings (A–E) cannot rotate and are fixed in space because of hindered single bond rotation. For example, the aromatic rings C and E have a chloro substituent which prevents these rings becoming coplanar with ring D. Similarly, rings A and B have phenol substituents which prevent them becoming coplanar.

The fixed conformation of the hexapeptide chain is important to vancomycin's unique mechanism of action, which involves targeting the cell wall's building blocks rather than a protein or a nucleic acid. To be specific, there is a pocket in the vancomycin structure into which the tail of the building block's pentapeptide moiety can fit. The pentapeptide is then held there by the formation of five hydrogen bonds between it and the hexapeptide chain of vancomycin (Fig. 19.54). Dimerization can now occur where a highly stable vancomycin dimer is bound to two tails. Because vancomycin is a large molecule, it caps the tails and acts as a steric shield,

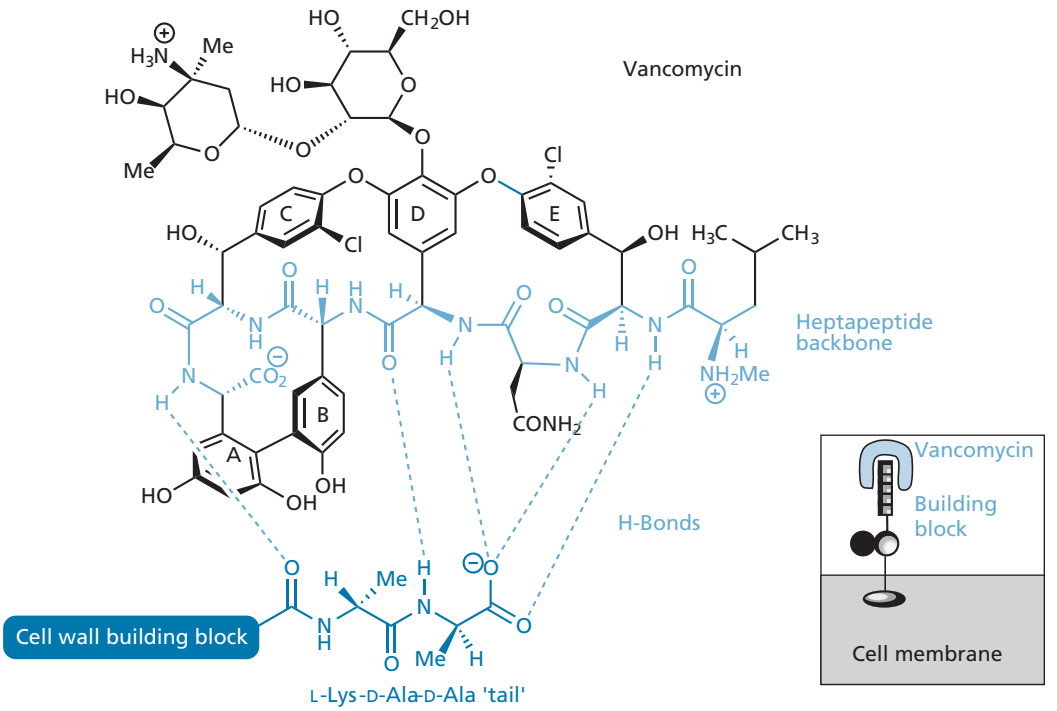


FIGURE 19.54 Vancomycin and its binding interactions to the L-Lys-D-Ala-D-Ala moiety.

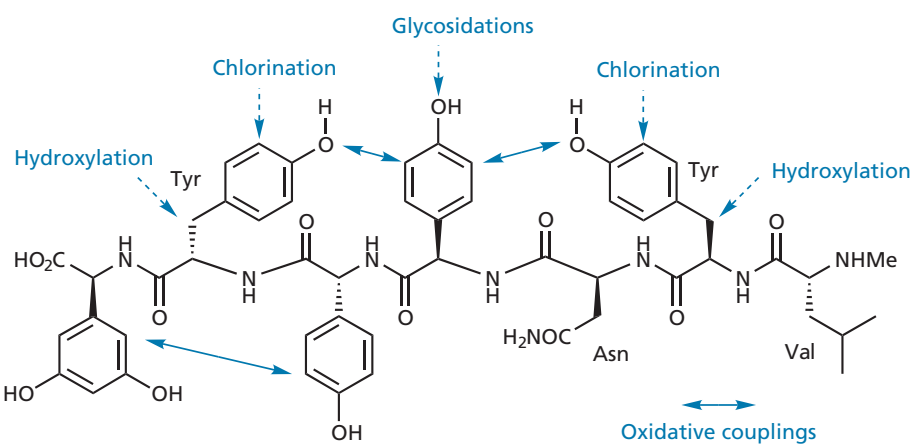


FIGURE 19.55 Reactions involved in the biosynthesis of vancomycin.

blocking access to the transglycosidase and transpeptidase enzymes (Fig. 19.56).

Dimerization occurs head to tail such that the heptapeptide chains of each vancomycin molecule interact through four hydrogen bonds (Fig. 19.57). The sugar and chloro-groups also play an important role in this dimerization, and activity drops if either of these groups is absent.

Because vancomycin is such a large molecule, it is unable to cross the outer cell membrane of Gram-negative bacteria and, consequently, lacks activity against those organisms. It is also unable to cross the inner cell membrane of Gram-positive bacteria, but this is not required as the construction of the cell wall takes place outside the cell membrane.

Bacterial resistance to vancomycin has been slow to develop, although some hospital strains of *S. aureus* were identified in 1996 which *do* show resistance (**VRSA**). Of particular concern was the appearance of **vancomycin-resistant enterococci** (**VRE**) in 1989. These are organisms that can cause life-threatening gut infections in patients whose immune system is weakened. Resistance in the latter organisms has arisen from a modification

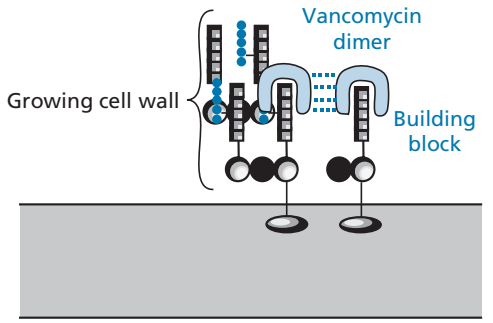


FIGURE 19.56 'Capping' of pentapeptide 'tails' by vancomycin.

of the cell wall precursors where the terminal D-alanine group in the pentapeptide chain has been replaced by D-lactic acid, resulting in a terminal ester link rather than an amide link (Fig. 19.58). This removes one of the NH groups involved in the hydrogen bonding interaction with vancomycin. It may not sound like much, but it is sufficient to weaken the binding affinity and make the antibiotic ineffective. The modified building block is still acceptable to the transglycosylase and transpeptidase enzymes. In the latter case, lactate acts as the leaving group rather than D-alanine.

Teicoplanin is a medication that contains five very similar structures which were isolated from a soil microorganism called *Actinoplanes teichomyceticus* and which differ only in the nature of a long alkyl substituent. One example is taicoplanin A_2–5 (Fig. 19.59). The teicoplanins belong to the vancomycin family but do not dimarize. The long alkyl chain plays an importart role is anchoring the antibiotic to the outer surface of the cell membrane where it is perfectly placed to interact with the building

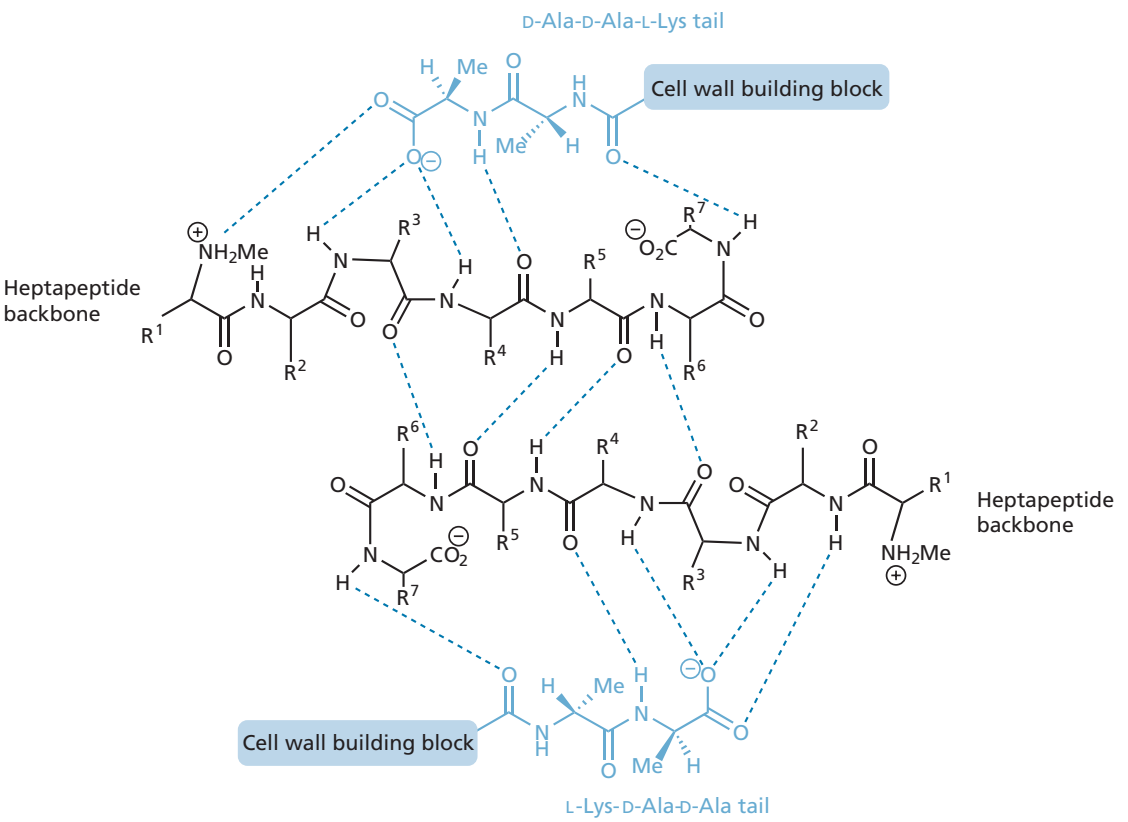


FIGURE 19.57 Dimerization of vancomycin. The dashed lines represent hydrogen bonds.

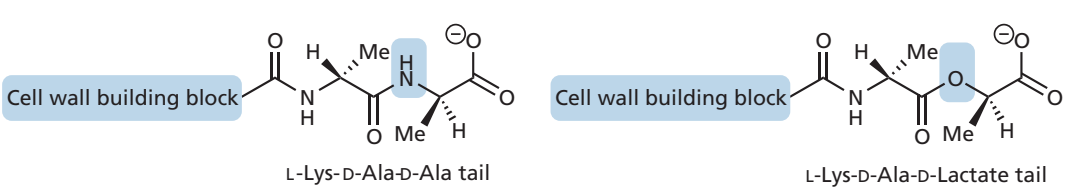


FIGURE 19.58 Modification of the pentapeptide chain leading to resistance.

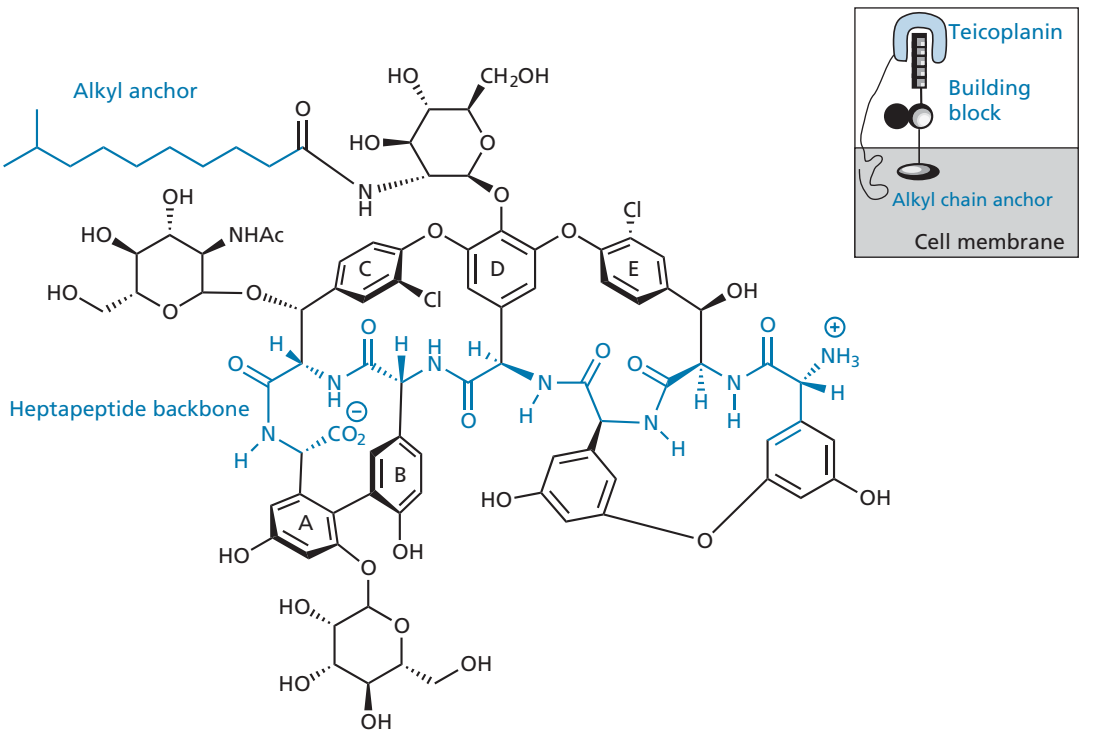


FIGURE 19.59 Teicoplanin A_2–5.

blocks for cell wall synthesis (Fig. 19.59). Teicoplanin is used clinically for the treatment of Gram-positive infections and is less toxic than vancomycin.

Another naturally occurring member of the vancomycin family is **eremomycin** (Fig. 19.60). A biphenyl hydrophobic 'tail' was added to act as an anchor, resulting in a compound (**LY 333328**), which is 1000 times more active than vancomycin. Further modifications involved removal of a tetrahydropyran ring to leave an alcohol group (R^4), modification of the hydrophobic tail (R^2) and addition of a side chain with a phosphate group (R^3), to give **telavancin**, which was approved in 2009.

Although the complexity of the glycopeptides is an advantage in their targeting and selectivity, it is a problem when it comes to synthesizing analogues. Therefore, work has been carried out to try and prepare simplified analogues of vancomycin which are easier to synthesize, yet retain the desired selectivity. Structures such as those shown in Fig. 19.61 have been prepared which are capable of binding to D-Ala -D-Ala and D-Ala-D-Lac. These now represent lead compounds for the development of future antibacterial agents.

There are another two mechanisms by which **glycopeptides** may have an antibacterial activity. Firstly, it is possible that glycopeptide dimers disrupt the cell membrane structure. This is supported by the fact that glycopeptide antibacterial agents enhance the activity of aminoglycosides by increasing their absorption through the cell membrane. Secondly, RNA synthesis is known to be disrupted in the presence of glycopeptides. The possibility of three different mechanisms of action explains why bacteria are slow to acquire resistance to the glycopeptides.

KEY POINTS

- β-Lactamase inhibitors are β-lactam structures that have negligible antibacterial activity but inhibit β-lactamases. They can be administered alongside penicillins to protect them from β-lactamases and to broaden their spectrum of activity.
- Carbapenems and monobactams are examples of other β-lactam structures with clinically useful antibacterial activity.
- Glycopeptides, such as vancomycin, bind to the building blocks for cell wall synthesis, preventing their incorporation into the cell wall. They also block the cross-linking reaction for those units already incorporated in the wall. The glycopeptides are the drugs of last resort against drug-resistant strains of bacteria.
- Bacitracin binds to and inhibits the carrier lipid responsible for carrying the cell wall components across the cell membrane.
- Cycloserine inhibits the synthesis of D-Ala-D-Ala.

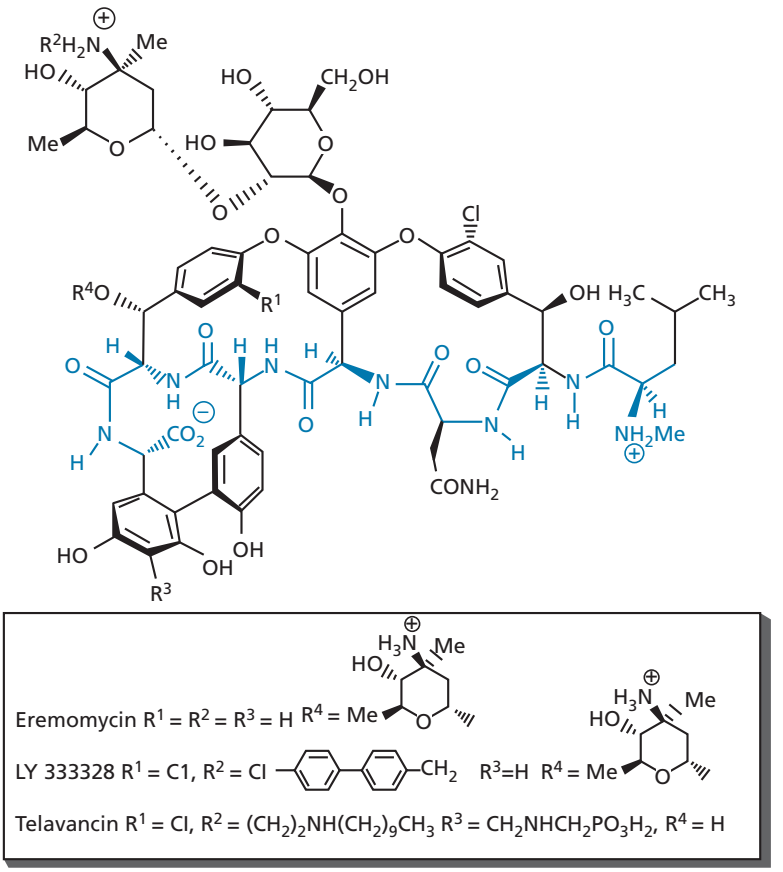


FIGURE 19.60 Eremomycin, LY 333328, and telavancin.

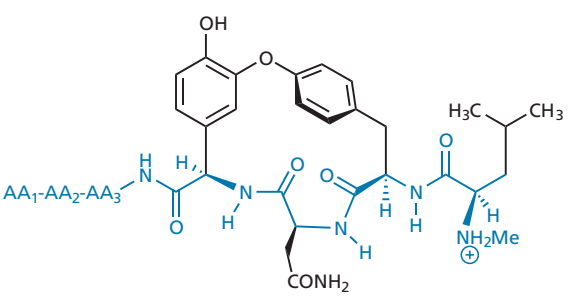


FIGURE 19.61 Simplified analogues of the glycopeptides.

19.6 Antibacterial agents which act on the plasma membrane structure

19.6.1 Valinomycin and gramicidin A

The peptides **valinomycin** and **gramicidin A** both act as ion-conducting antibiotics (ionophores) and allow the uncontrolled movement of ions across the cell membrane. These agents are described in section 10.6.

19.6.2 Polymyxin B

The polypeptide antibiotic **polymyxin B** (Fig. 19.62) derives from a soil bacterium called *Bacillus polymyxa*. It also operates within the cell membrane and shows a selective toxicity for bacterial cells over animal cells. This appears to be related to the ability of the compound to bind selectively to the different plasma membranes. The mechanism of this selectivity is not fully understood. Polymyxin B acts like valinomycin (section 10.6.2), but it causes the leakage of small molecules such as nucleosides from the cell.

19.6.3 Killer nanotubes

Work is in progress to design cyclic peptides which will self-assemble in the cell membranes of bacteria to form tubules that have been labelled as killer **nanotubes** (section 10.6.1).

BOX 19.12 Clinical aspects of cycloserine, bacitracin, and vancomycin

D-Cycloserine is administered orally in combination with other drugs to treat tuberculosis that is resistant to first-line drugs. **Bacitracin** is used alongside **polymyxin B sulphate** for the topical treatment of skin infections. The same preparation is also used for the topical treatment of eye infections caused by *P. aeruginosa*. **Neomycin sulphate** and bacitracin are used together for the topical treatment of skin infections as a cream or dusting powder.

Vancomycin and **teicoplanin** are bactericidal and are active against aerobic and anaerobic Gram-positive bacteria, including MRSA. Vancomycin is not absorbed orally and is administered by intravenous injection every 12 hours for the prophylaxis and treatment of endocarditis and other serious infections caused by Gram-positive cocci. Vancomycin is also given orally to treat gut infections caused by *Clostridium difficile*. This organism may appear following the use of broad-spectrum antibiotics and produces harmful toxins. Vancomycin has also been used in eye drops. Teicoplanin can be administered once daily and is used for potentially serious Gram-positive infections, including endocarditis, dialysis-associated peritonitis, and serious infections caused by *S. aureus*. It is also used as a prophylactic in endocarditis and orthopaedic surgery. Telavancin was approved in 2009 for the treatment of skin infections which include MRSA.

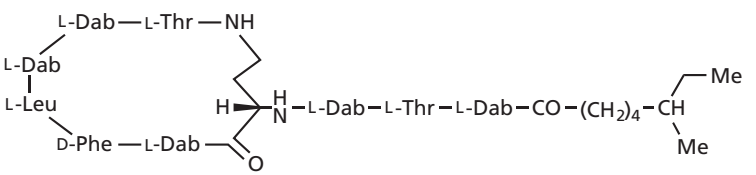


FIGURE 19.62 Polymyxin B (Dab = α,γ-Diaminobutyric acid with peptide link through the α-amino group).

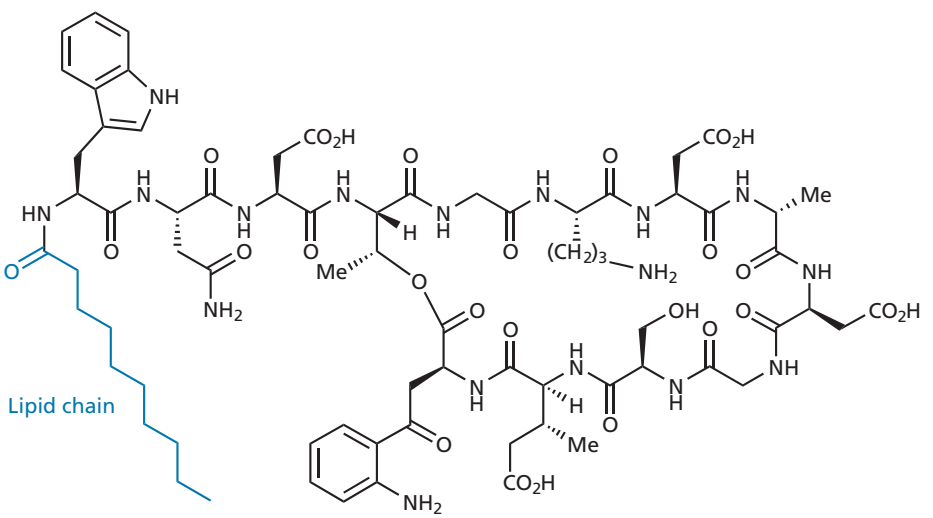


FIGURE 19.63 Daptomycin.

19.6.4 Cyclic lipopeptides

Daptomycin (Fig. 19.63) is a member of a new class of antibacterial agents called the cyclic lipopeptides. It is a natural product derived from a bacterial strain *Streptomyces roseosporus*, and works by disrupting multiple functions of the bacterial cell membrane. The lipid portion of the molecule is derived from decanoic acid and the yield of product obtained is increased if decanoic acid is added to the fermentation medium.

KEY POINTS

- Ionophores act on the plasma membrane and result in the uncontrolled movement of ions across the cell membrane, leading to cell death.

BOX 19.13 Clinical aspects of drugs acting on the plasma membrane

Valinomycin and **gramicidin A** show no selective toxicity for bacterial cells over mammalian cells and are therefore useless as systemic therapeutic agents. However, gramicidin is present as a minor constituent in some topical applications.

Polymyxin B is injected intramuscularly and is useful against *Pseudomonas* strains that are resistant to other antibacterial agents. It can be used topically for the treatment of minor skin infections and has good activity against Gram-negative bacteria. It is less effective against Gram-positive bacteria, as it is difficult for such a big molecule to pass through the thicker cell wall. It is used in combination with **bacitracin** for the treatment of eye and skin infections, or with **dexamethasone** and **neomycin** for the treatment of eye infections. The ear-drop preparation **Otosporin** contains **hydrocortisone** and polymyxin B to treat ear infections and inflammation.

Daptomycin was approved in 2003 for the treatment of Gram-positive infections. It is administered by intravenous infusion and has a spectrum of activity similar to vancomycin. In order to guard against the development of drug resistance, the drug is held in reserve for skin and soft tissue infections caused by drug-resistant Gram-positive bacteria, such as MRSA. It can be administered alongside other antibacterial agents for mixed infections involving Gram-positive bacteria, Gram-negative bacteria, and some anaerobes.

- Polymyxin B operates selectively on the plasma membrane of bacteria and causes the uncontrolled movement of small molecules across the membrane.
- Cyclic peptides are being designed which will self-assemble to form nanotubes in the cell membranes of bacteria.
- Cyclic lipopeptides are a new class of antibiotic.

19.7 Antibacterial agents which impair protein synthesis: translation

The agents described in this section all inhibit protein synthesis by binding to ribosomes and inhibiting different stages of the translation process (Fig. 19.64). Selective toxicity is due to either different diffusion rates through the cell barriers of bacterial versus mammalian cells or to a difference between the ribosomal target structures. The bacterial ribosome is a 70S particle (see section 6.2.2) made up of a 30S subunit and a 50S subunit. The 30S subunit binds messenger RNA (mRNA) and initiates protein synthesis. The 50S subunit combines with the 30S subunit-mRNA complex to form a ribosome, then binds aminoacyl transfer RNA (tRNA) and catalyses the building of the protein chain. There are two main binding sites for the tRNA molecules. The peptidyl site (P-site) binds the tRNA bearing the peptide chain. The acceptor aminoacyl site (A-site) binds the tRNA bearing the next amino acid, to which the peptide chain will be transferred (see also section 6.2.2). The ribosomes of eukaryotic cells are bigger (80S), consisting of 60S and 40S subunits. They are sufficiently different in structure from prokaryotic ribosomes that it is possible for some drugs to distinguish between them.

19.7.1 Aminoglycosides

Streptomycin (Fig. 19.65) was isolated from the soil microorganism *Streptomyces griseus* in 1944 and is an

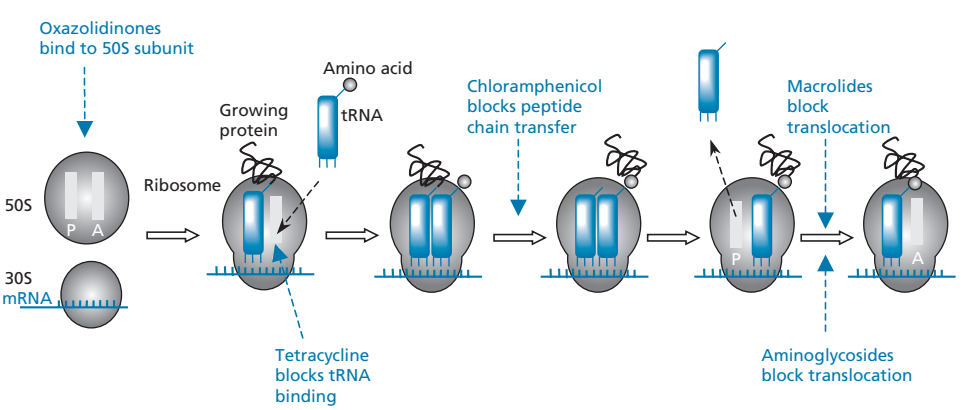
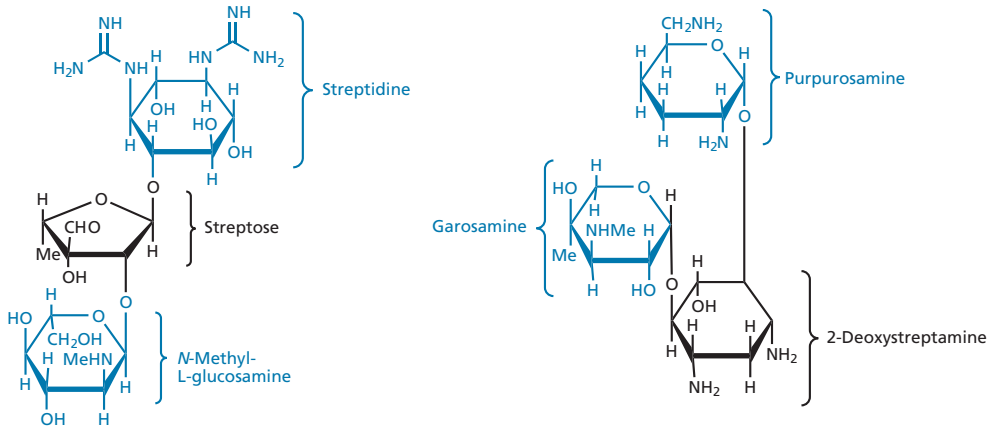


FIGURE 19.64 Stages at which antibacterial agents inhibit translation.



Streptomycin (from *Streptomyces griseus*) Gentamicin C1a

FIGURE 19.65 Aminoglycosides.

BOX 19.14 Clinical aspects of aminoglycosides

Aminoglycosides are fast acting, but they can also cause ear and kidney problems if the dose levels are not carefully controlled. They are effective in the treatment of infections caused by aerobic Gram-negative bacteria, including *P. aeruginosa*. Indeed, they used to be the only compounds effective against that organism. Some Gram-negative bacteria are resistant to aminoglycosides due mainly to enzymes which catalyse reactions such as *O*-phosphorylations, *O*-adenylations (addition of an adenine group), and *N*-acylations. Resistance can also occur from alterations of the ribosomes such that they bind aminoglycosides less strongly or by less efficient uptake mechanisms. Because the aminoglycosides are polar in nature, they have to be injected. They are also unable to cross the blood–brain barrier efficiently and so they cannot be used for the treatment of meningitis unless they are injected directly into the central nervous system (CNS). The activity of aminoglycosides is increased if they are administered with agents which disrupt cell wall synthesis, as this increases uptake into the cell. However, bacteriostatic agents should not be taken with aminoglycosides, because these inhibit the energy-dependent uptake process by which the aminoglycosides cross the cell membrane.

Streptomycin was the first effective agent used against tuberculosis. However, resistance soon developed and a multidrug therapy involving streptomycin, **isoniazid**, and ***para*-aminosalicylic acid** was used until the early 1970s. At that point, **rifampicin** became available, allowing different multidrug therapies to be developed. Streptomycin is now rarely used for the treatment of tuberculosis, unless there is a known resistance to isoniazid, in which case it is administered by intramuscular injection. Streptomycin is still used to treat enterococcal endocarditis and as an adjunct to **doxycycline** in brucellosis.

Gentamycin is administered by intramuscular or slow intravenous injection for the treatment of a number of infections, including septicaemia; neonatal sepsis; CNS infections (including meningitis); biliary tract infections; acute pyelonephritis or prostatis, endocarditis; and pneumonia in hospital patients. It can be used topically in drops for the treatment of eye and ear infections.

example of an aminoglycoside—a carbohydrate structure which includes basic amine groups. Streptomycin was the next most important antibiotic to be discovered after penicillin and a variety of other aminoglycosides have been subsequently isolated from various organisms, for example **gentamicin C1a** (Fig. 19.65). The aminoglycosides work best in slightly alkaline conditions. At pH 7.4, they have a positive charge that is beneficial to activity by aiding absorption through the outer membrane of Gram-negative bacteria. An ionic interaction takes place with various negatively charged groups on the outer surface of the cell membrane which displaces magnesium and calcium ions. These ions normally act as bridges between lipopolysaccharides, and their displacement results in rearrangement of cell membrane components to produce pores through which an aminoglycoside can pass. The drug then crosses the cell membrane by an energy-dependent process and is trapped inside the cell where it accumulates to relatively high concentrations.

Binding to bacterial ribosomes now takes place to inhibit protein synthesis. The binding is specifically to the 30S ribosomal subunit and prevents the movement of the ribosome along mRNA so that the triplet code on mRNA can no longer be read. In some cases, protein synthesis is terminated and the shortened proteins end up in the cell membrane. This can lead to a further increase in cell permeability, resulting in an even greater uptake of the drug. Aminoglycosides are bactericidal rather than bacteriostatic and it is thought that their activity may be due to their effects both on the ribosomes and the outer cell membrane.

Because the ribosomes in human cells are different in structure from those in bacterial cells, they have a much lower binding affinity for the aminoglycosides, which explains the selectivity of these drugs.

19.7.2 Tetracyclines

The tetracyclines are bacteriostatic antibiotics which have a broad spectrum of activity and are the most widely prescribed form of antibiotic after penicillins. They are also capable of attacking the malarial parasite. One of the best known tetracyclines is **chlortetracycline** (**aureomycin**) (Fig. 19.66), which was isolated in 1948 from a mud-growing microorganism in Missouri called *Streptomyces aureofaciens*—so-called because of its golden colour. Further tetracyclines, such as **tetracycline** and **doxycycline** (Fig. 19.66), have been synthesized or discovered.

The tetracyclines inhibit protein synthesis by binding to the 30S subunit of ribosomes and preventing aminoacyl-tRNA from binding. This stops the further addition of amino acids to the growing protein chain. Protein release is also inhibited.

In the case of Gram-negative bacteria, tetracyclines cross the outer membrane by passive diffusion through the porins. Passage across the inner membrane is dependent on a pH gradient, which suggests that a proton-driven carrier is involved. Selectivity is due to the ability of bacterial cells to concentrate these agents faster than human cells. This is fortunate because tetracyclines are capable of inhibiting protein synthesis in mammalian cells—particularly in mitochondria.

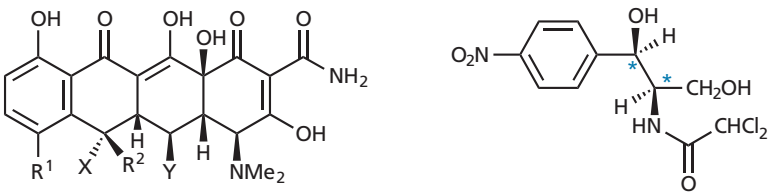
BOX 19.15 Clinical aspects of tetracyclines and chloramphenicol

The tetracyclines are broad-spectrum antibiotics with activity against both Gram-positive and Gram-negative bacteria. Commonly used tetracyclines in the clinic are **tetracycline**, **demeclocycline**, **doxycycline**, **lymecycline**, **minocycline**, and **oxytetracycline**. The use of chlortetracycline has decreased over the years because it kills the intestinal flora that produce vitamin K. However, it is still administered alongside tetracycline and demeclocycline in the preparation **Deteclo**.

In general, the tetracyclines can be divided into short-lasting compounds, such as chlortetracycline, an intermediate group of compounds which includes demeclocycline, and longer-acting compounds which include doxycycline. Minocycline has the broadest spectrum of activity. The tetracyclines were used originally for many types of respiratory infections, but have been largely replaced by β-lactams because of the problems of resistance. However, they are still the agents of choice for the treatment of Lyme disease, rickettsia, and infections caused by *Chlamydia* species. They are also used to treat acne and a variety of different infections including respiratory and genital infections. Doxycycline has been found to be useful for the treatment and prophylaxis of malaria, and is cheaper than other antimalarial agents. One drawback is the possibility of skin hypersensitivity to sunlight. The drug can also be used for the treatment of a variety of diseases including syphilis, sinusitis, oral herpes simplex, and acne. It is a possible agent for the treatment or prophylaxis of anthrax.

Tetracyclines should be avoided for young children and pregnant mothers as they can bind to developing teeth and bone leading to tooth discolouration. Resistance to tetracyclines can arise through several mechanisms. Some organisms have effective efflux mechanisms which pump the drug back out of the cell. Resistance can also arise from alterations in the bacterial ribosomes such that they have lower affinity for the agents.

Chloramphenicol is a potent broad-spectrum antibiotic and, in some regions of the world, it is the drug of choice for the treatment of typhoid when more expensive drugs cannot be afforded. It can also be used in severe bacterial infections which are insensitive to other antibacterial agents and is widely used for eye infections. It can also be used for ear infections, but the preparation can cause hypersensitivity reactions in about 10% of patients. The drug should only be used in these restricted scenarios as it is quite toxic, especially to bone marrow. The drug is metabolized inadequately in babies leading to a combination of symptoms described as the **gray baby syndrome**, which can be fatal. In adults, the drug undergoes a phase II conjugation reaction to form a glucuronic acid conjugate (section 11.5.5), which is excreted. This reaction fails to take place efficiently in newborn babies and so the drug levels increase to toxic levels. Bacteria with resistance to the drug contain an enzyme called **chloramphenicol acetyltransferase**, which catalyses the acylation of the hydroxyl groups.



Chlortetracycline (Aureomycin) (R^1 = Cl, R^2 = Me, X = OH, Y = H)
Tetracycline (R^1 = H, R^2 = Me, X = OH, Y = H)
Doxycycline (Vibramycin) (R^1 = H, R^2 = Me, X = H, Y = OH)
Demeclocycline (R^1 = Cl, R^2 = H, X = OH, Y = H)

Chloramphenicol

FIGURE 19.66 Tetracyclines and chloramphenicol. The asterisks indicate asymmetric centres.

Widespread resistance to tetracyclines has occurred, caused partly by the use of tetracyclines to cure animal infections and as a food additive to promote the growth of newborn animals.

Test your understanding and practise your molecular modelling with Exercise 19.5.

19.7.3 Chloramphenicol

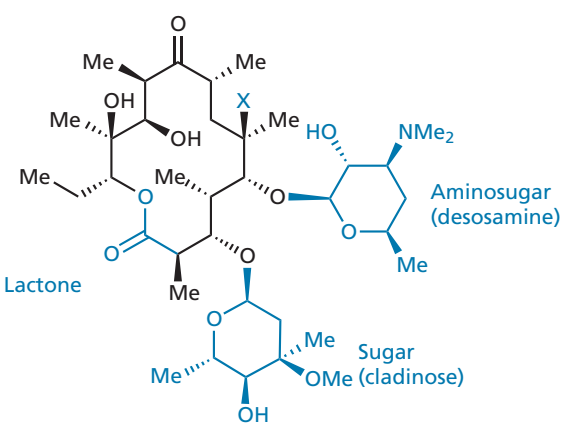
Chloramphenicol (Fig. 19.66) was originally isolated from a microorganism called *Streptomyces venezuela* found in a field near Caracas, Venezuela. It is now prepared synthetically and has two asymmetric centres. Only the *R*,*R*-isomer is active.

Chloramphenicol binds to the 50S subunit of ribosomes and appears to act by inhibiting the movement of ribosomes along mRNA, probably by inhibiting the peptidyl transferase reaction by which the peptide chain is extended. Since it binds to the same region as macrolides and **lincosamides**, these drugs cannot be used in combination. The nitro group and both alcohol groups are involved in binding interactions. The dichloroacetamide group is also important, but can be replaced by other electronegative groups. Chloramphenicol is quite toxic and the nitro substituent is thought to be responsible for this.

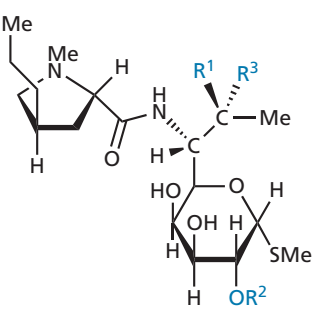
19.7.4 Macrolides

Macrolides are bacteriostatic agents. The best-known example of this class of compounds is **erythromycin**—a metabolite isolated in 1952 from the soil microorganism *Streptomyces erythreus* found in the Philippines, and one of the safest antibiotics in clinical use. The structure (Fig. 19.67) consists of a 14-membered macrocyclic lactone ring with a sugar and an amino sugar attached. The sugar residues are important for activity.

Erythromycin acts by binding to the 50S subunit of bacterial ribosomes to inhibit translocation, but other mechanisms of action also appear likely. Because erythromycin and chloramphenicol bind to the same



Erythromycin; X = OH
Clarithromycin; X = OMe



Lincomycin R^1=R^2=H, R^3=OH
Clindamycin R^1=Cl, R^2=R^3=H
Clindamycin phosphate R^1=Cl, R^2=PO_3^{2-}, R^3=H

FIGURE 19.67 Macrolides and lincosamides.

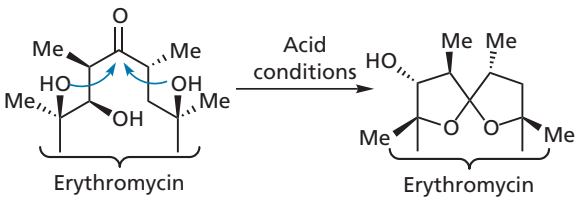


FIGURE 19.68 Intramolecular ketal formation in erythromycin.

region of the ribosome, they should not be administered together as they will compete with each other and be less effective.

Erythromycin is unstable to stomach acids, but can be taken orally in a tablet form. The formulation of the tablet involves a coating that is designed to protect the tablet during its passage through the stomach, but which is soluble once it reaches the intestines (enterosoluble). The acid sensitivity of erythromycin is due to the presence of a ketone and two alcohol groups which are set up for the acid-catalysed intramolecular formation of a ketal (Fig. 19.68). One way of preventing this is to protect the hydroxyl groups. For example, **clarithromycin** is a methoxy analogue of erythromycin which is more stable to gastric juices and has improved oral absorption. Another method of increasing acid stability is to increase the size of the macrocycle to a 16-membered ring.

Azithromycin (Fig. 19.69) contains a 15-membered macrocycle where an *N*-methyl group has been incorporated into the macrocycle. It is one of the world's best-selling drugs. **Telithromycin** (Fig. 19.69) is a semi-synthetic derivative of erythromycin and reached the European market in 2001. The cladinose sugar in erythromycin has been replaced with a keto-group and a carbamate ring has been fused to the macrocyclic ring. The two hydroxyl groups that cause the intramolecular ketal formation in erythromycin have been masked, one as a methoxy group and the other as part of the carbamate ring.

19.7.5 Lincosamides

The lincosamide antibiotics (Fig. 19.67) have similar antibacterial properties to the macrolides and act in the same fashion. **Lincomycin** was the first of these agents and was isolated in 1962 from a soil organism called *Streptomyces lincolnensis* found near Lincoln, Nebraska. Chemical modification led to the clinically useful **clindamycin** with increased activity.

19.7.6 Streptogramins

Pritinamycin is a mixture of macrolactone structures obtained from *Streptomyces pristinaespiralis*. Two of the components (**quinupristin** and **dalfopristin**) have been isolated. These agents bind to different regions of the bacterial ribosome's 50S subunit form a complex. It is found that binding of dalfopristin increases the binding affinity for quinupristin, and so the two agents act in synergy with each other. Quinupristin inhibits peptide chain elongation, while dalfopristin interferes with the transfer of the peptide chain from one tRNA to the next.

19.7.7 Oxazolidinones

The **oxazolidinones** are a new class of synthetic antibacterial agents discovered in recent years. They inhibit protein synthesis at a much earlier stage than previous agents, and, consequently, do not suffer the same resistance problems. Before protein synthesis can start, a 70S ribosome has to be formed by the combination of a 30S ribosome with a 50S ribosome. The oxazolidinones bind to the 50S ribosome and prevent this from happening. As a result, translation cannot even start. Other agents that inhibit protein synthesis do so during the translation process itself (Fig. 19.64). **Linezolid** (Fig. 19.70) was the first of this class of compounds to reach the market in 2000, and by 2010, it was netting sales of £716 million

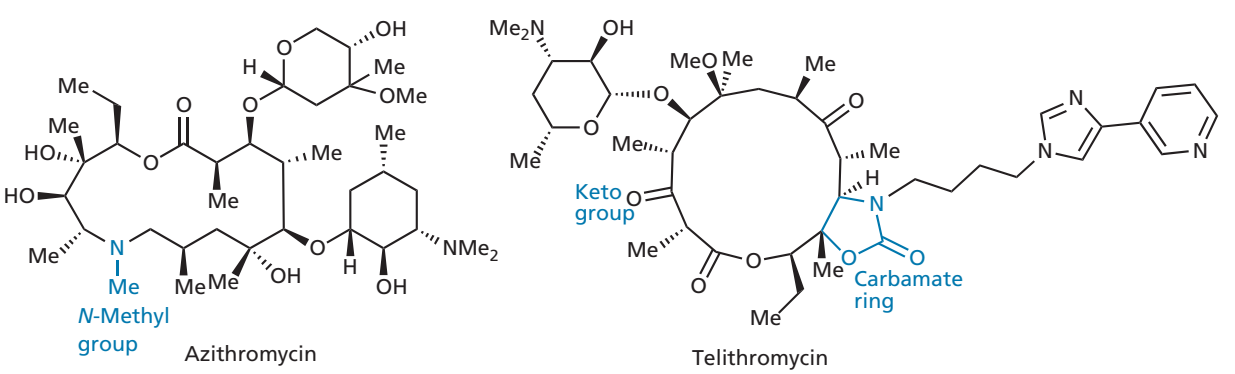


FIGURE 19.69 Azithromycin and telithromycin.

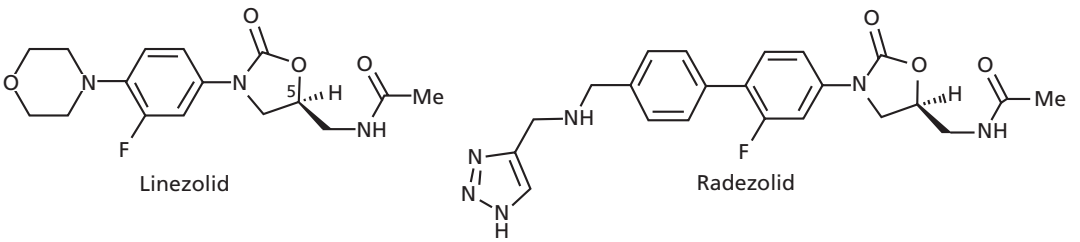


FIGURE 19.70 Oxazolidinones.

BOX 19.16 Clinical aspects of macrolides, lincosamides, streptogramins, and oxazolidinones

Macrolides

Erythromycin has an antibacterial spectrum that is similar to penicillins and can be used as an alternative to penicillins for those patients having penicillin allergies. It has been used against penicillin-resistant staphylococci, but newer penicillins are now preferred for these infections owing to increased resistance against erythromycin. It is very useful for the treatment of respiratory infections, including whooping cough and Legionnaires' disease. It can also be used to treat syphilis and diphtheria, as well as oral and skin infections. Topically, it can be used for the treatment of acne. **Clarithromycin** has slightly greater activity than erythromycin, with fewer gastrointestinal side effects. Therefore, it is often prescribed instead of erythromycin. Clarithromycin is one of the drugs used in the treatment of ulcers caused by the presence of *Helicobacter pylori* (section 25.4). **Azithromycin** is slightly less active than erythromycin against Gram-positive infections, but is more active against Gram-negative infections, including *H. influenza*—against which erythromycin shows poor activity. Azithromycin can also be used for the treatment of Lyme disease. **Telithromycin** has a similar spectrum of activity to other macrolides. It should only be used for specified infections such as pneumonia, tonsillitis, and sinusitis.

Resistance to macrolides is due to effective efflux mechanisms which pump the drug back out the cell. The ribosomal target site may also change in character such that binding is weakened. Enzyme-catalysed modifications can also occur. Recently there has been research into finding novel macrolides which can be effective against respiratory infections due to resistant strains of *S. pneumoniae*, as well as the organism *H. influenza*.

Lincosamides

Clindamycin can be taken orally and is active against Gram-positive cocci, including streptococci and penicillin-resistant staphylococci. It is active against peripheral infections involving the anaerobic *B. fragilis*, and is recommended for the treatment of joint and bone infections caused by staphylococci. It is also used topically for the treatment of acne.

Streptogramins

Pritinamycin has been used orally in the treatment of Gram-positive cocci infections, including MRSA. **Quinupristin** and **dalfopristin** are used intravenously in combination (**Synercid**). At present, these agents are reserved for life-threatening Gram-positive infections for which there are no alternative therapies; for example hospital-acquired pneumonia, skin and soft tissue infections, and infections caused by vancomycin-resistant *Enterococcus faecium*.

Oxazolidinones

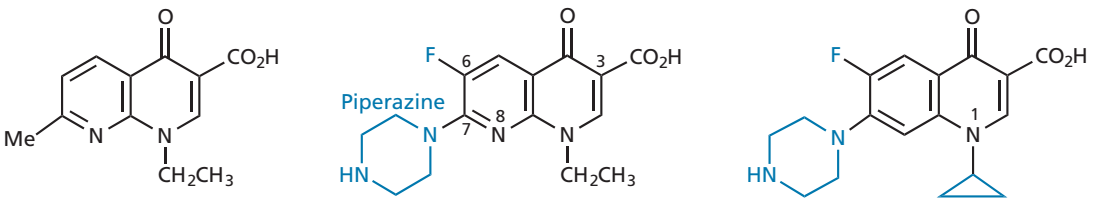
The oxazolidinones have a broad spectrum of activity and are active against bacterial strains which have acquired resistance to other antibacterial agents acting against protein synthesis. **Linezolid** has good activity against most clinically important Gram-positive bacteria, including MRSA. It can also be taken orally with 100% uptake from the gastrointestinal tract. Unfortunately, there is a high level of side effects related to its use and, as it is a bacteriostatic agent, there is a greater risk of bacterial resistance developing.

per year. X-ray crystallographic studies have revealed how the structure binds to the ribosome, and that has allowed the development of analogues which bind more strongly. **Radezolid** is one such structure which binds 10,000 times more strongly as a result of extra binding interactions (*extension strategy*, section 13.3.2). It is currently undergoing clinical trials.

19.8 Agents that act on nucleic acid transcription and replication

19.8.1 Quinolones and fluoroquinolones

The quinolone and **fluoroquinolone** antibacterial agents are particularly useful in the treatment of urinary tract



Nalidixic acid Enoxacin Ciprofloxacin

FIGURE 19.71 Quinolones and fluoroquinolones.

infections and infections which prove resistant to the more established antibacterial agents.

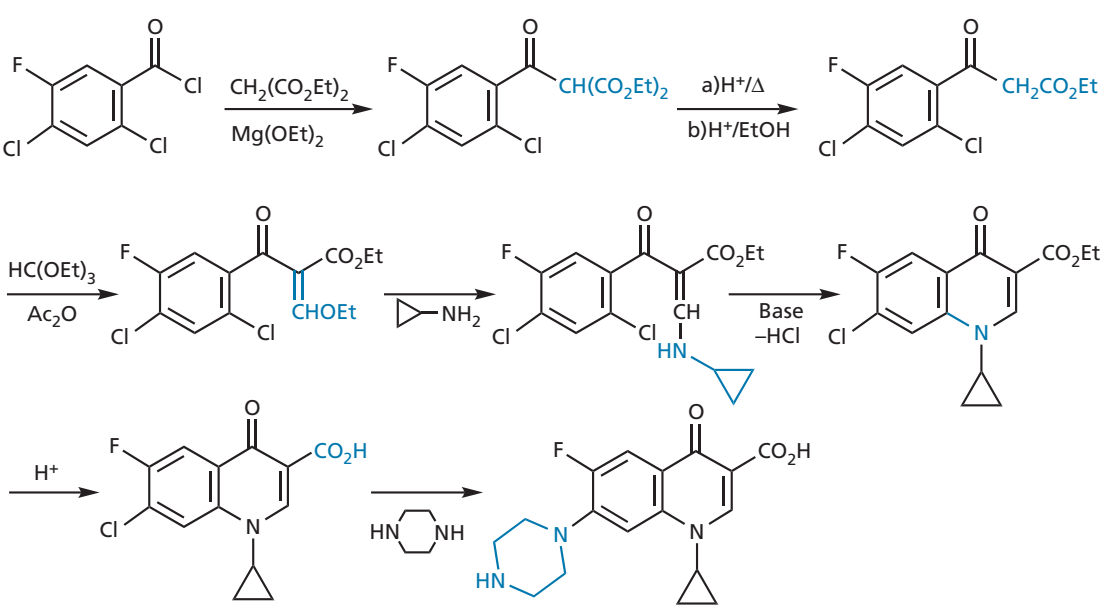
Nalidixic acid (Fig. 19.71), synthesized in 1962, was the first therapeutically useful agent in this class of compounds. Various analogues were synthesized but offered no great advantage. However, a breakthrough was made in the 1980s with the development of **enoxacin** (Fig. 19.71), which showed improved broad-spectrum activity. The development of enoxacin was based on the discovery that a single fluorine atom at position 6 greatly increased both activity and cellular uptake. A basic substituent, such as a piperazinyl ring at position 7, was also beneficial for a variety of pharmacokinetic reasons due to the ability of the basic substituent to form a zwitterion with the carboxylic acid group at position 3.

The introduction of a cyclopropyl substituent at position 1 further increased broad-spectrum activity, while replacement of the nitrogen at position 8 with carbon reduced adverse reactions and increased activity against *S. aureus*. This led to **ciprofloxacin** (Fig. 19.71 and Box 19.17), the most active of the fluoroquinolones against Gram-negative bacteria.

The quinolones and fluoroquinolones inhibit the replication and transcription of bacterial DNA by stabilizing

BOX 19.17 Synthesis of ciprofloxacin

The synthesis of ciprofloxacin is a seven-stage route and is applicable to a wide range of fluoroquinolones. It involves the construction of the 'right-hand' pyridone ring onto the fluoro-substituted aromatic ring. The cyclopropyl substituent is incorporated just before ring closure and the piperazinyl substituent is added at the final stage of the synthesis.



Synthesis of ciprofloxacin.

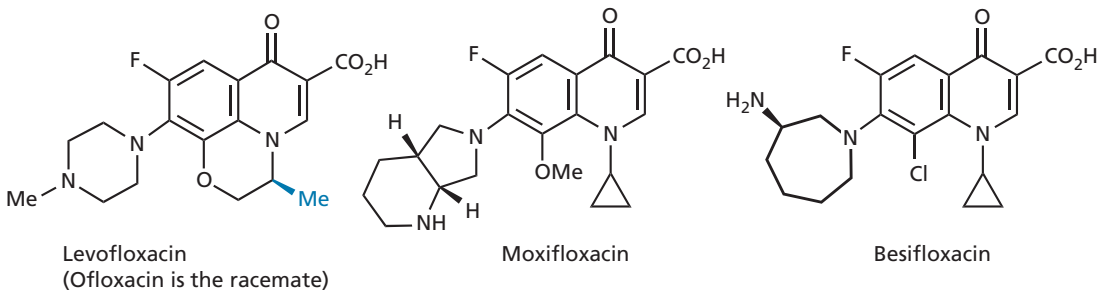


FIGURE 19.72 Third- and fourth-generation fluoroquinolones.

the complex formed between DNA and topoisomerases (section 9.2). In Gram-positive bacteria, the stabilized complexes are between DNA and **topoisomerase IV**, with the drugs showing a 1000-fold selectivity for the bacterial enzyme over the corresponding enzyme in human cells. In Gram-negative bacteria, the main target for fluoroquinolones is the complex between DNA and a topoisomerase II enzyme called **DNA gyrase**. It has the same role as topoisomerase IV in reverse and is required when the DNA double helix is being supercoiled after replication and transcription.

A large number of fluoroquinolones have now been synthesized. Those agents having good activity all have a similar bicyclic ring system, which includes a pyridone ring and a carboxylic acid at position 3. A problem with first- and second-generation fluoroquinolones is that they generally show only moderate activity against *S. aureus,* followed by rapidly developing drug resistance. Furthermore, only marginal activity is shown against anaerobes and *S. pneumoniae*. Third- and fourth-generation fluoroquinolones, such as **ofloxacin**, **levofloxacin**, **moxifloxacin**, and **besifloxacin** (Fig. 19.72) began to be developed in the early 1990s to tackle these issues. Ofloxacin has an asymmetric centre and is sold as a racemic mixture of both enantiomers, one of which is active and one of which is not. Levofloxacin is the active enantiomer of oflaxacin and is twice as active as the racemate.

19.8.2 Aminoacridines

Aminoacridine agents, such as the yellow-coloured **proflavine**, are topical antibacterial agents which were used particularly during World War II to treat deep surface wounds. The best agents are completely ionized at pH 7 and they interact directly with bacterial DNA by intercalation (section 9.1). Despite the success of these drugs

BOX 19.18 Clinical aspects of quinolones and fluoroquinolones

Nalidixic acid is active against Gram-negative bacteria and is useful in the short-term therapy of uncomplicated urinary tract infections. It can be taken orally but, unfortunately, bacteria can develop a rapid resistance to it. **Enoxacin** has a greatly increased spectrum of activity against Gram-negative and Gram-positive bacteria. It also shows improved oral absorption, tissue distribution, and metabolic stability, as well as an improvement in the level and spectrum of activity, particularly against Gram-negative bacteria, such as *P. aeruginosa*. **Ciprofloxacillin** is used in the treatment of a large range of infections involving the urinary, respiratory, and gastrointestinal tracts (e.g. travellers' diarrhoea), as well as infections of skin, bone, and joints. It is also used for gonorrhoea and septicaemia, and as part of a cocktail of drugs for anthrax. It has been claimed that ciprofloxacin may be the most active broad-spectrum antibacterial agent on the market. In contrast to nalidixic acid, resistance to the fluoroquinolones is slow to appear, but, when it does appear, it is mainly due to efflux mechanisms which pump the drug back out of the cell. Less common resistance mechanisms include mutations to the topoisomerase enzymes which reduce their affinity to the agents, and alteration of porins in the outer membrane of Gram-negative organisms to limit access.

Third-generation fluoroquinolones show improved activity against *S. pneumoniae*, while maintaining activity against enterobacteria. **Ofloxacin** is administered orally or by intravenous infusion to treat septicaemia, gonorrhoea, and infections of the urinary tract, lower respiratory tract, skin, and soft tissue. **Levofloxacin** has a greater activity against pneumococci than ciprofloxacillin and is a second-line treatment for community-acquired pneumonia. It is also used for acute sinusitis, chronic bronchitis, urinary tract infections, skin infections, and soft tissue infections. **Moxifloxacin** also has greater activity against pneumococci than ciprofloxacin. It is used to treat sinusitis and is a second-line treatment of community-acquired pneumonia. **Besifloxacin** is a fourth-generation fluoroquinolone approved in 2009.

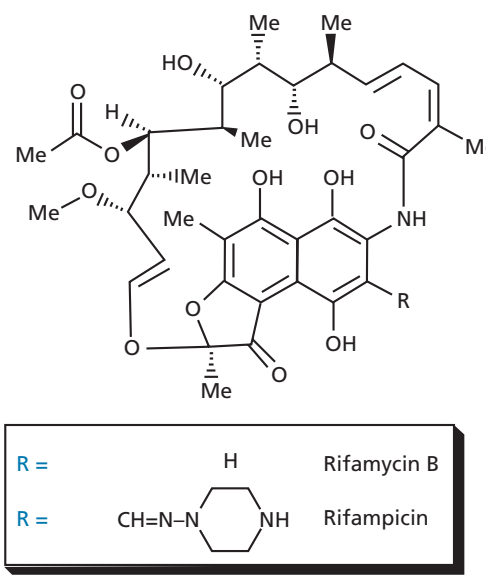
as topical agents, they are not suitable for the treatment of systemic bacterial infections because they are toxic to host cells.

19.8.3 Rifamycins

Rifampicin (Fig. 19.73) is a semi-synthetic rifamycin made from **rifamycin B**—an antibiotic which was isolated from *Streptomyces mediterranei* in 1957. It inhibits Gram-positive bacteria and works by binding non-covalently to **DNA-dependent RNA polymerase** and inhibiting the start of RNA synthesis. The DNA-dependent RNA polymerases in eukaryotic cells are unaffected because the drug binds to a peptide chain not present in the mammalian RNA polymerase. It is, therefore, highly selective. The flat naphthalene ring and several of the hydroxyl groups are essential for activity and the molecule exists as a zwitterion, giving it good solubility both in lipids and aqueous acid. **Rifaximin** is another semi-synthetic analogue that was approved in 2004 for the treatment of diarrhoea and *E. coli* infection.

19.8.4 Nitroimidazoles and nitrofurantoin

Metronidazole (Fig. 19.73) is a nitroimidazole structure which was introduced in 1959 as an anti-protozoal agent, but began to be used as an antibacterial agent in the 1970s. The nitro group is reduced when the drug



Rifaximin

Metronidazole

Nitrofurantoin

Methenamine

Fusidic acid

Isoniazid

Pyrazinamide

(+)Ethambutol

FIGURE 19.73 Miscellaneous agents.

enters the bacterial cell, which lowers the concentration of metronidazole within the cell and sets up a concentration gradient down which more drug can flow. The reduction mechanism also proves toxic to the cell as free radicals are formed which act on DNA. **Nitrofurantoin** also undergoes reduction within bacterial cells to form radical species that act on DNA.

19.8.5 Inhibitors of bacterial RNA polymerase

A recent addition to the arsenal of clinically useful antibiotics is **fidaxomicin** (Fig. 19.74), which is a natural product obtained from a *Dactylosporangium* Gram-positive bacterial strain. The agent is a macrocycle and was approved in 2011 as a narrow spectrum bactericidal agent for the treatment of *C. difficile* infections in the gastrointestinal tract. It inhibits transcription in *C. difficile* by inhibiting RNA polymerase, and has a minimal effect on other gut flora.

19.9 Miscellaneous agents

A variety of miscellaneous agents are shown in Fig. 19.73. **Methenamine** is used to treat urinary tract infections where it degrades in acid conditions to give formaldehyde as the active agent (section 14.6.6). **Fusidic acid** is a steroid structure derived from the fungus *Fusidium coccineum* and is used as a topical antibacterial agent. **Isoniazid** is the most widely used drug for the treatment of tuberculosis. It acts by inhibiting the synthetic pathways leading to mycolic acid, an important constituent of mycobacterial cell walls. It is activated in bacterial cells by a catalase-peroxidase enzyme. Resistant strains of tuberculosis block the action of this enzyme. **Ethambutol** and **pyrazinamide** are synthetic compounds which are both front-line drugs in the treatment of tuberculosis. Ethambutol inhibits **arabinosyl transferase** enzymes that are involved in the biosynthesis of the mycobacterial cell wall.

KEY POINTS

- Aminoglycosides, tetracyclines, chloramphenicol, streptogramins, lincosamides, and macrolides inhibit protein synthesis by binding to the bacterial ribosomes involved in the translation process.
- Resistance can arise from a variety of mechanisms, such as drug efflux, altered binding affinity of the ribosome, altered membrane permeability, and metabolic reactions.
- Oxazolidinones prevent the formation of the 70S ribosome by binding to the 50S subunit.
- Quinolones and fluoroquinolones inhibit topoisomerase enzymes, resulting in inhibition of replication and transcription.
- Aminoacridines are useful topical antibacterial agents which can intercalate with bacterial DNA and hinder replication and transcription.
- Rifamycins inhibit the enzyme RNA polymerase and prevent RNA synthesis. In turn, this prevents protein synthesis. Rifampicin is used to treat tuberculosis and staphylococcus infections. Fidaxomicin is a macrocycle which also targets RNA polymerase.
- Nitroimidazoles are used against infections caused by protozoa and anaerobic bacteria.

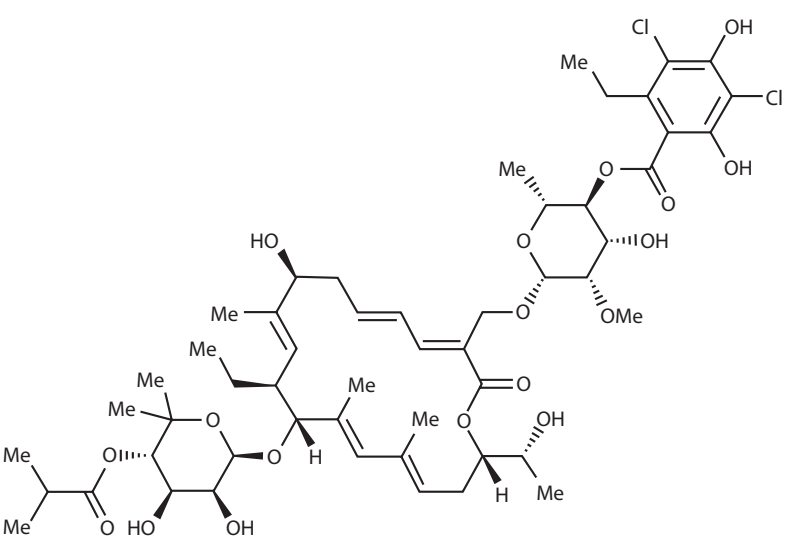


FIGURE 19.74 Fidaxomicin.

BOX 19.19 Clinical aspects of rifamycins and miscellaneous agents

Rifampicin is bactericidal and is used mainly in the treatment of tuberculosis and staphylococci infections that resist penicillin. It is used in combination with dapsone in treating leprosy and is also used for the treatment of brucellosis, legionnaires' disease, and serious staphylococcal infection. It is a very useful antibiotic, showing a high degree of selectivity against bacterial cells over mammalian cells. Unfortunately, it is also expensive, which discourages its use against a wider range of infections. Rifampicin is a key component of any anti-tuberculosis regimen, but it poses a special problem when treating tuberculosis in AIDS patients, as it enhances the activity of the cytochrome P450 enzyme family (CYP3A). These enzymes metabolize the HIV protease inhibitors used in HIV therapy, thus lowering their effectiveness. Increased cytochrome P450 activity also decreases the effect of oral anticoagulants, oral contraceptives, and barbiturates.

Metronidazole has good activity in treating infections caused by anaerobic bacteria and protozoa, including difficult-to-treat organisms, such as *B. fragilis* and *C. difficile*. It is well distributed round the body and crosses the blood–brain barrier, so it can be used for the treatment of brain abscesses and other central nervous system infections involving anaerobic bacteria. Metronitrazole is used for the treatment of leg ulcers, bacterial vaginosis, pelvic inflammatory disease, and can also be used as an alternative to penicillins for oral infections, including tooth abscesses. It is administered with amoxicillin (or with tetracycline and bismuth) in the treatment of gastric ulcers involving *H. pylori* (section 25.4). The drug is effective against *Giardia* infections derived from polluted water supplies—a common hazard when visiting the third world. Finally, nitroimidazoles, such as metronidazole, are commonly combined with cephalosporins or aminoglycosides to treat infections involving both aerobic and anaerobic organisms. Resistance is rare, though not of the question. **Nitrofurantoin** is used to treat uncomplicated urinary tract infections.

Methenamine can be used to treat urinary tract infections, but only if the urine is acidic and the infection is in the lower urinary tract. It can be used as a prophylactic, and as a treatment for chronic and recurrent lower urinary tract infections.

Fusidic acid is a topical antibacterial agent that is used in eye drops and skin creams. It can penetrate intact and damaged skin, so it is useful for the treatment of boils. It has also been used to eradicate MRSA colonies carried in the nasal passages of hospital patients and health workers.

Isoniazid is the most widely used drug for the treatment of tuberculosis and is part of a four-drug cocktail which is the first choice treatment for the initial phase of the disease.

19.10 Drug resistance

Medicinal chemists are still actively seeking new and improved antibacterial agents to combat the worrying ability of bacteria to acquire resistance to current drugs. For example, 60% of *S. pneumoniae* strains are resistant to β-lactams, and 60% of *S. aureus* strains are resistant to **methicillin**. The last resort in treating *S. aureus* infections is **vancomycin**, but resistance is also beginning to appear to that antibiotic. Some strains of *E. faecalis* appearing in urinary and wound infections are resistant to all known antibiotics and are untreatable. If antibiotic resistance continues to grow, medicine could be plunged back to the 1930s. Indeed, many of today's advanced surgical procedures would become too risky to carry out due to the risks of infection. Old diseases are already making a comeback. For example, a new antibiotic-resistant strain of tuberculosis [**multidrug-resistant TB** (**MDRTB**)] appeared in New York and took 4 years and $10 million to bring under control. These strains were resistant to two of the front-line drugs used against tuberculosis (isoniazid and rifamycin), and had various levels of resistance against another two (streptomycin and ethambutol). Other examples of bacterial strains acquiring resistance include penicillin-resistant meningococci and pneumococci in South Africa, penicillin-resistant gonococci in Asia and Africa, ampicillin-resistant *H. influenza* in the USA and Europe, and chloramphenicol-resistant meningococci in France and Southeast Asia. Resistance to trimethoprim in some of the developing nations has meant that the drug has become ineffective as a treatment for dysentery.

Drug resistance can arise because of a variety of factors described in section 19.5.1.5, but the cell must have the necessary genetic information. This information can be obtained by mutation or by the transfer of genes between cells.

19.10.1 Drug resistance by mutation

Bacteria multiply at such a rapid rate that there is always a chance that a mutation will render a bacterial cell resistant to a particular agent. This feature has been known for a long time and is the reason why patients should complete a full course of antibacterial treatment

even though their symptoms may have disappeared well before the end of the course. If this rule is adhered to, the vast majority of invading bacterial cells will be wiped out, leaving the body's own defences to mop-up any isolated survivors or resistant cells. If the treatment is stopped too soon, however, then the body's defences struggle to cope with the survivors. Any isolated resistant cell is then given the chance to multiply, resulting in a new infection which will, of course, be completely resistant to the original drug. This was a major factor in the appearance of MDRTB.

Mutations occur naturally and randomly, and do not require the presence of a drug. Indeed, it is likely that a drug-resistant cell is present in a bacterial population even before the drug is encountered. This was demonstrated with the identification of **streptomycin-resistant** cells from old cultures of *E. coli*, which had been freeze-dried to prevent multiplication before the introduction of streptomycin into medicine.

19.10.2 Drug resistance by genetic transfer

A second way in which bacterial cells can acquire drug resistance is by gaining that resistance from another bacterial cell. This occurs because it is possible for genetic information to be passed on from one bacterial cell to another. There are two main methods by which this can take place—**transduction** and **conjugation**.

In transduction, small segments of genetic information known as **plasmids** are transferred by means of bacterial viruses (**bacteriophages**) which leave the resistant cell and infect a non-resistant cell. If the plasmid contains the gene required for drug resistance, then the recipient cell will be able to use that information and gain resistance. For example, the genetic information required to synthesize β-**lactamases** can be passed on in this way, rendering bacteria resistant to penicillins. The problem is particularly prevalent in hospitals where over 90% of staphylococcal infections are currently resistant to antibiotics such as penicillin, erythromycin, and tetracycline. It may seem odd that hospitals should be a source of drug-resistant strains of bacteria. In fact, they are the perfect breeding ground. Drugs commonly used in hospitals are present in the air in trace amounts. It has been shown that breathing in these trace amounts kills sensitive bacteria in the nose and allows the nostrils to act as a breeding ground for resistant strains.

In conjugation, bacterial cells pass genetic material directly to each other. This is a method used mainly by Gram-negative, rod-shaped bacteria in the colon, and involves two cells building a connecting bridge of sex pili through which the genetic information can pass.

19.10.3 Other factors affecting drug resistance

The more useful a drug is, the more it will be prescribed, and the greater the possibilities of resistant bacterial strains emerging. The original penicillins were used widely in human medicine, but were also used commonly in veterinary medicine. Antibacterial agents have also been used in animal feeding to increase animal weight and this, more than anything else, has resulted in drug-resistant bacterial strains. It is sobering to think that many of the original bacterial strains which were treated so dramatically with penicillin V or penicillin G are now resistant to those early penicillins. In contrast, these two drugs are still highly effective antibacterial agents in poorer, developing African nations, where the use (and abuse) of the drugs has been far less widespread.

The ease with which different bacteria acquire resistance varies. For example, *S. aureus* is notorious for its ability to acquire drug resistance owing to the ease with which it can undergo transduction. However, the microorganism responsible for syphilis seems incapable of acquiring resistance and is still susceptible to the original drugs used against it.

19.10.4 The way ahead

The ability of bacteria to gain resistance to drugs is an ever-present challenge to the medicinal chemist and it is important to continue designing new antibacterial agents. Identifying potential new targets is essential in this never-ending battle. The sequencing of genomes and a study of the proteins present in bacterial cells promises to give more detailed understanding of the molecular details of infectious agents leading to the identification of new drug targets. For example, *Mycobacterium tuberculosis*—the causative agent of tuberculosis—has a complex cell wall where three types of polymers are attached to peptidoglycan. The detailed mechanisms by which these polymers are synthesized and incorporated into the cell wall are being investigated to identify new targets for antibacterial drugs which will disrupt the cell wall structure.

It is also beginning to be appreciated that the drugs with the least susceptibility to resistance are those with several different modes of action. Therefore, designing drugs which act on a number of different targets, rather than one specific target, is more likely to be successful.

Examples of new targets include kinase enzymes. There has already been success in designing kinase inhibitors as anticancer agents (section 21.6.2) and several research groups are now looking at agents that might prove to be selective inhibitors of bacterial kinases. Other potential targets are the enzymes known as **aminoacyl**

tRNA synthetases. These enzymes are an ancient group of enzymes responsible for attaching amino acids to tRNA. Because they are ancient, there is a considerable sequence divergence between the bacterial and human enzymes, making selective inhibition possible. **Isoleucyl tRNA synthetase** is one such enzyme which is known to be inhibited by **mupirocin** (Fig. 19.75)—a clinically useful antibiotic isolated from *Pseudomonas fluorescens* with activity against MRSA. Mupirocin is used as a topical agent for skin infections and has also been used to combat the transmission of *S. aureus* within hospitals by treating the nasal passages of patients and hospital staff. Unfortunately, the widespread use of the agent for this purpose has led to strains of *S. aureus* with increasing resistance to the drug. Research is now being carried out to find novel inhibitors for a different aminoacyl tRNA synthetase present in *S. aureus*, namely **tyrosine tRNA synthetase**. The strategy of targeting aminoacyl tRNA synthetases is also proving fruitful in the search for novel antifungal agents. **Tavaborole** inhibits **leucine tRNA synthetase** and is undergoing clinical trials for the treatment of fungal nail infections.

Another potential approach in countering resistance is to modify antibiotics such that they gain resistance to the mechanisms of resistance used against them! For example, **kanamycin** is an aminoglycoside which is no longer used because resistant bacteria can phosphorylate one of the hydroxyl groups present (Fig. 19.76).

An active analogue has been synthesized which replaces the susceptible alcohol with a ketone (Fig. 19.77). This ketone is in equilibrium with the hydrated gem-diol. When phosphorylation occurs on the diol, the phosphate group thus formed acts as a good leaving group and the ketone is regenerated. *In vitro* tests showed that this agent was active against strains of bacteria which are resistant to kanamycin.

Another approach is to design molecules with an inbuilt self-destruct mechanism. One of the problems with antibiotics in medicine or veterinary practice is that much of the active antibiotic is excreted, giving bacteria in the environment the opportunity to gain resistance. This problem could be reduced by incorporating a self-destruct mechanism which kicks in once the antibiotic is excreted. For example, work has been carried out on a cephalosporin containing a protected hydrazine group (Fig. 19.78). The protecting group concerned is *ortho*-nitrobenzylcarbamate which is susceptible to light. Once the antibiotic is excreted and exposed to light, the protecting group is lost, allowing the nucleophilic hydrazine moiety to react with the β-lactam ring and deactivate the molecule. This works *in vitro* but has still be tested *in vivo*.

Recent research into drug combinations has shown that there can be a beneficial effect on antibacterial activity *in vitro* if one administers an antibacterial drug with another drug, even if the other drug has no antibacterial activity itself. For example, a small dose of the tetracycline agent **minocycline** showed better activity than expected when it was administered along with the anti-diarrhoeal drug **loperamide** (Box 24.3). Further studies

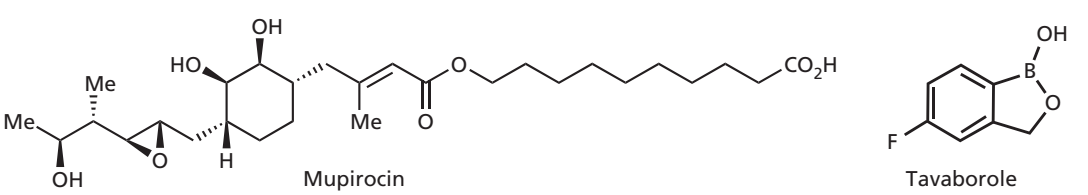


FIGURE 19.75 Inhibitors of aminoacyl tRNA synthetases.

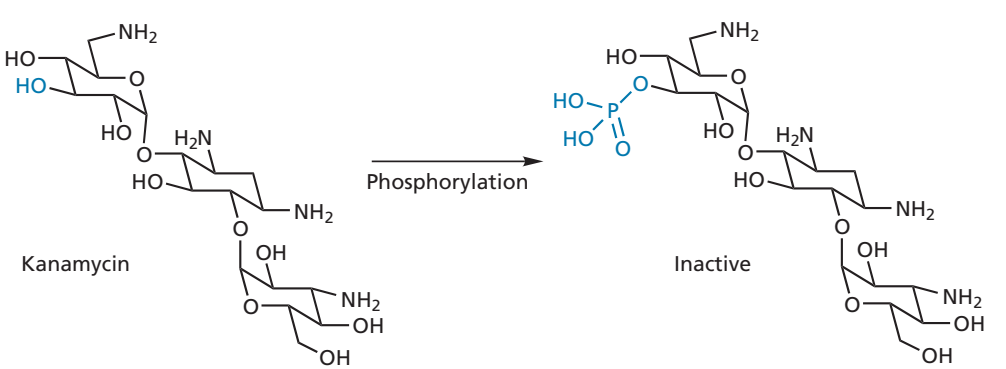


FIGURE 19.76 The phosphorylation reaction causing resistance to kanamycin.

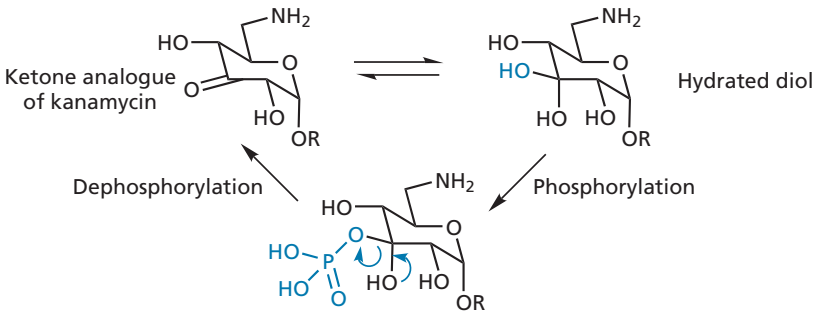


FIGURE 19.77 Analogue of kanamycin which is resistant to phosphorylation.

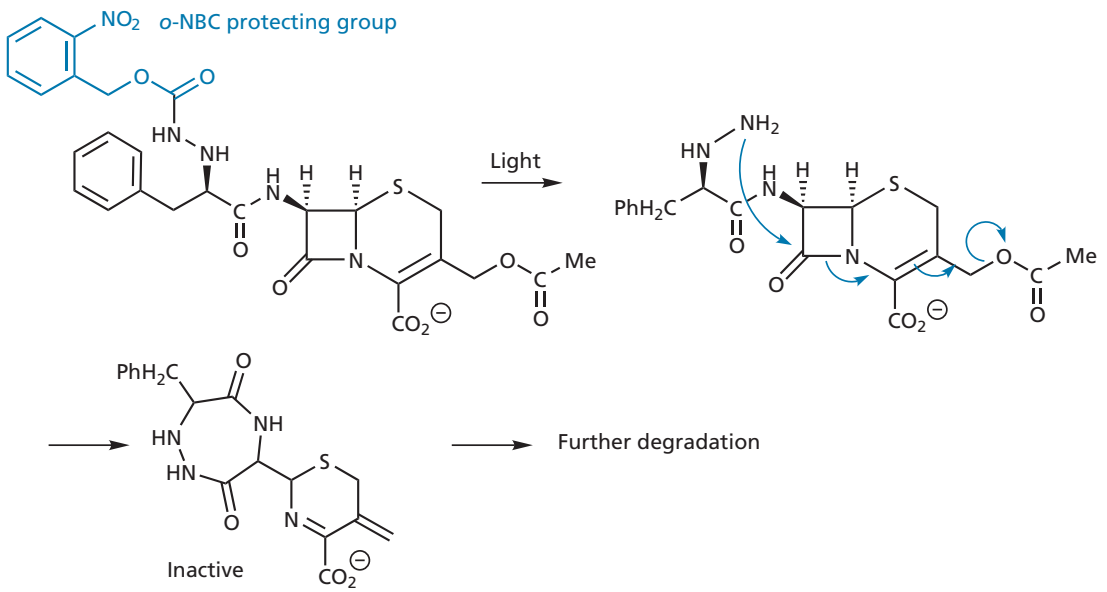


FIGURE 19.78 Self-destruct mechanism.

BOX 19.20 Organoarsenicals as antiparasitic drugs

The first effective antimicrobial drug to be synthesized was the organoarsenical, **salvarsan** (section 19.1). In the late 1940s, another organoarsenical called **melarsoprol** was introduced into medicine and is the first-choice drug for the treatment of trypanosomiasis and sleeping sickness. This is despite the fact that it has to be injected and can kill 1 in 20 patients treated with it (Fig. 1).

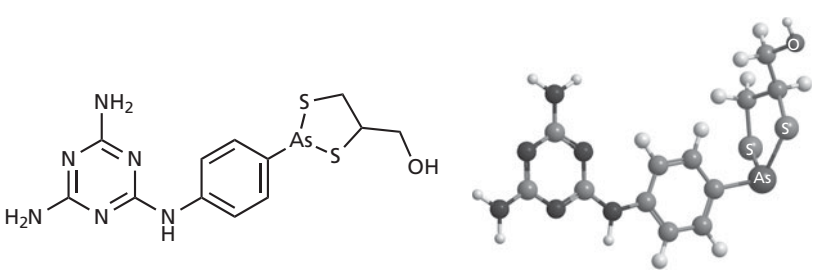


FIGURE 1 Melarsoprol.

(*Continued*)

Box 19.20 Organoarsenicals as antiparasitic drugs (*Continued*)

One of the mechanisms by which melarsoprol might act is through a reaction with the cysteine residues of enzymes involved in glycolysis (Fig. 2). The blocking of glycolysis leads to a loss of cell motility and eventual cell death. Other mechanisms of action have been proposed.

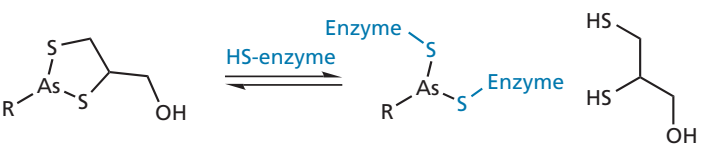


FIGURE 2 Mechanism of action of melarsoprol.

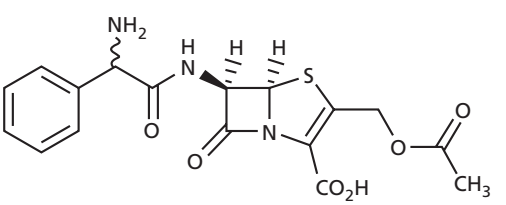
are needed to see if this effect occurs *in vivo*, but it might be another way of tackling drug resistance.

KEY POINTS

- Bacterial strains vary in their ability to gain resistance to antibacterial drugs. *Staphylococcus aureus* is quick to gain antibacterial resistance. The MRSA strain is a *S. aureus* strain that is resistant to most antibacterials, including methicillin.
- Vancomycin is the antibacterial agent of last resort in the treatment of resistant bacterial strains.
- There are many mechanisms by which bacteria can acquire resistance against antibacterial agents, but they all result from a change in the cell's genetic make-up.
- Drug resistance can result from mutation of a cell's genetic information or from transfer of genetic information from one cell to another. Genetic information can be transferred from one cell to another by transduction or conjugation.
- Care has to be taken to use antibacterial agents in a responsible manner to reduce the chances of resistance developing.
- It is important to identify new targets which can be used for the design of novel antibacterial agents.

QUESTIONS

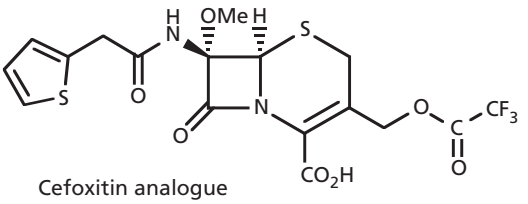
1. How would you convert penicillin G to 6-aminopenicillanic acid (6-APA) using chemical reagents? Suggest how you would make ampicillin from 6-APA.
2. Penicillin is produced biosynthetically from cysteine and valine. If the biosynthetic pathway could accept different amino acids, what sort of penicillin analogues might be formed if valine was replaced by alanine, phenylalanine, glycine, or lysine? What sort of penicillin analogue might be formed if cysteine was replaced by serine? (See Appendix 1 for amino acid structures.)
3. Referring to Question 2, why do you think penicillin analogues like this are not formed during the fermentation process?
4. The activity of sulphonamides is decreased if they are taken at the same time as procaine. Suggest why this might be the case.
5. Discuss whether you think the following penicillin analogue would be a useful antibacterial agent.



Penicillin analogue

6. Explain what effect replacing the methoxy groups on methicillin with ethoxy groups might have on the properties of the agent.
7. What effect might the bicyclic ring system of cephalosporins have on their chemical and biological properties compared with the bicyclic ring system of penicillins, and why?

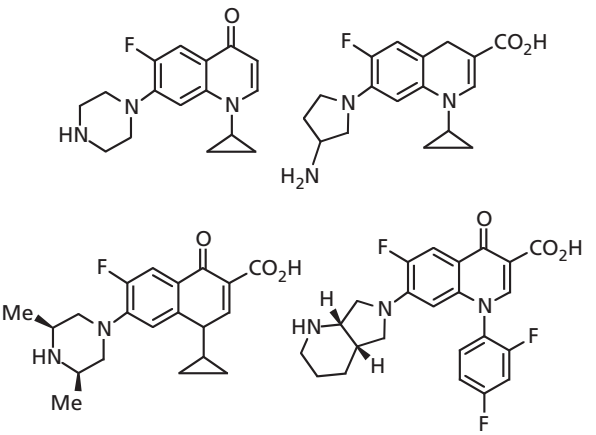
8. The following structure is an analogue of cefoxitin. What sort of properties do you think it might have compared to cefoxitin itself?



Cefoxitin analogue

9. Show the mechanism by which the prodrug bacampicillin (Box 19.7) is converted to ampicillin. What are the by-products?

10. Which of the following structures would you expect to have the best antibacterial activity?



11. Devise a synthesis for the structure chosen in Question 10.

FURTHER READING

Armstrong, D. and Cohen, J. (eds) (1999) *Infectious Diseases*, Section 7. Mosby, London.

Broadwith, P. (2010) Rousing sleeping sickness research. *Chemistry World* May: 23.

Coates, A., Hu, Y., Bax, R., and Page, C. (2002) The future challenges facing the development of new antimicrobial drugs. *Nature Reviews Drug Discovery* **1**, 895–910.

Evans, J. (1998) TB: Know your enemy. *Chemistry in Britain* November, 38–42.

Hook, V. (1997) Superbugs step up the pace. *Chemistry in Britain* May, 34–35.

King, A. (2012) Making light work. *Chemistry World* April, 52–55.

Kirkpatrick, P., Raja, A., LaBonte, J., and Lebbos, J. (2003) Daptomycin. *Nature Reviews Drug Discovery*, **2**, 943–944.

Mendell, G., Bennett, J. E., and Dolin, R. (eds) (2000) *Mendell, Douglas and Bennett's Principles and Practice of Infectious Diseases*, 5th edn , Vols 1 and 2. Churchill Livingstone, Edinburgh.

Sansom, C. (2012) The latent threat of tuberculosis. *Chemistry World*, September, 48–51.

Škedelj, V., Tomašić, T., Mašič, L.P., and Zega, A. (2011) ATP-Binding site of bacterial enzymes as a target for antibacterial drug design. *Journal of Medicinal Chemistry* **54**, 915–929.

Lactams and other agents acting on cell walls

Axelsen, P. H. and Li, D. (1998) A rational strategy for enhancing the affinity of vancomycin towards depsipeptide ligands. *Bioorganic and Medicinal Chemistry* **6**, 877–881.

Nicolaou, K. C., Boddy, C. N., Bräse, S., and Winssinger, N. (1999) Chemistry, biology, and medicine of the glycopeptide antibiotics. *Angewandte Chemie International Edition* **38**, 2096–2152.

Agents acting on the cell membrane

Mann, J. (2001) Killer nanotubes. *Chemistry in Britain* November, 22.

Linezolid

Ford, C. (2001) First of a kind. *Chemistry in Britain* March, 22–24.

Genin, M. J., Hutchinson, D. K., Allwine, D. A., Hester, J. B., Emmert, D. E., Garmon, S. A., *et al.* (1998) N-C-linked (azolylphenyl)oxazolidinones. *Journal of Medicinal Chemistry* **41**, 5144–5147.

Quinolones and other agents acting on nucleic acids

Andriole, V. T. (Ed.) (1998) *The Quinolones*. 2nd edn. Academic Press, New York.

Saunders, J. (ed.) (2000) Quinolones as anti-bacterial DNA gyrase inhibitors. *Top Drugs: Top Synthetic Routes*. Oxford University Press, Oxford.

Agents acting against protein synthesis

Agouridas, C., Denis, A., Auger, J. M., Benedetti, Y., Bonnefoy, A., Bretin, F. *et al.* (1998) Synthesis and antibacterial activity of ketolides. *Journal of Medicinal Chemistry* **41**, 4080–4100.

Titles for general further reading are listed on p. 763.

20 Antiviral agents

20.1 Viruses and viral diseases

Viruses are non-cellular, infectious agents which take over a host cell in order to survive and multiply. There are many different viruses capable of infecting bacterial, plant, or animal cells, with more than 400 known to infect humans. Those capable of being transmitted to humans from animals or insects can be particularly dangerous and belong to a class of diseases defined as **zoonoses**. Consequently, both human and veterinary medicine play important roles in the control of such diseases.

Viruses can be transmitted in a variety of ways. Those responsible for diseases such as influenza (flu), chicken pox, measles, mumps, viral pneumonia, rubella, and smallpox can be transmitted through the air by an infected host sneezing or coughing. Other viruses can be transmitted by means of arthropods or ticks, leading to diseases such as Colorado tick fever and yellow fever. Some viruses are unable to survive for long outside the host and are transmitted through physical contact. The viruses responsible for AIDS, cold sores, the common cold, genital herpes, certain leukaemias, and rabies are examples of this kind. Finally, food- or water-borne viruses can lead to hepatitis A and E, poliomyelitis, and viral gastroenteritis.

Historically, viral infections have proved devastating to human populations. It has been suggested that smallpox was responsible for the major epidemics which weakened the Roman Empire during the periods AD 165–180 and AD 251–266. Smallpox was also responsible for the decimation of indigenous tribes in both North and South America during European colonization. In some areas, it is estimated that 90% of the population died from the disease. Various flu epidemics and pandemics have proved devastating. The number of deaths worldwide due to the flu pandemic of 1918–1919 is estimated to be over 20 million—far larger than the number killed by military action during World War I. Finally, it is estimated that 30 million people have died as a result of HIV infection since the 1980s.

The African continent has its fair share of lethal viruses, including Ebola and the virus responsible for Lassa fever. In the past, viral diseases such as these occurred in isolated communities and were easily contained. Nowadays, with cheap and readily available air travel, tourists are able to visit remote areas, thus increasing the chances of rare or new viral diseases spreading around the world. Therefore, it is important that world health authorities monitor potential risks and take speedy, appropriate action when required. The outbreak of **severe acute respiratory syndrome** (SARS) in the Far East during 2003 could have had a devastating effect worldwide if it had been ignored. Fortunately, the world community acted swiftly and the disease was brought under control relatively quickly. Nevertheless, the SARS outbreak serves as a timely warning of how dangerous viral infections can be. Scientists have warned of a nightmare scenario involving the possible evolution of a 'supervirus'. Such an agent would have a transmission mode and infection rate equivalent to flu, but a much higher mortality rate. There are already lethal viruses which can be spread rapidly and have a high mortality rate. Fortunately, the latency period between infection and detectable symptoms is short and so it is possible to contain the outbreak, especially if it is in isolated communities. If such viral infections evolved such that the latency period increased to that of AIDS, they could result in devastating pandemics equivalent to the plagues of the Middle Ages.

Considering the potential devastation that viruses can wreak on society, there are fears that terrorists might one day try to release lethal viral strains on civilian populations. This has been termed **bioterrorism**. To date, no terrorist group has carried out such an action, but it would be wrong to ignore the risk.

It is clear that research into effective antiviral drugs is a major priority in medicinal chemistry.

20.2 Structure of viruses

At their simplest, viruses can be viewed as protein packages transmitting foreign nucleic acid between host cells. The type of nucleic acid present depends on the virus

concerned. All viruses contain one or more molecules of either RNA or DNA, but not both. They can, therefore, be defined as **RNA** or **DNA viruses**. Most RNA viruses contain single-stranded RNA (**ssRNA**), but some viruses contain double-stranded RNA (**dsRNA**). If the base sequence of the RNA strand is identical to viral mRNA, it is called the positive (+) strand. If it is complementary, it is called the negative (–) strand. Most DNA viruses contain double-stranded DNA (**dsDNA**), but a small number contain single-stranded DNA (**ssDNA**). The size of the nucleic acid varies widely, with the smallest viral genomes coding for 3–4 proteins and the largest coding for over 100 proteins.

The viral nucleic acid is contained and protected within a protein coat called the **capsid**. Capsids are usually made up of protein subunits called **protomers** which are generated in the host cell and can interact spontaneously to form the capsid in a process called **self-assembly**. Once the capsid contains the viral nucleic acid, the whole assembly is known as the **nucleocapsid**. In some viruses, the nucleocapsid may contain viral enzymes which are crucial to its replication in the host cell. For example, the flu virus contains an enzyme called **RNA-dependent RNA polymerase** within its nucleocapsid (Fig. 20.1).

Additional membranous layers of carbohydrates and lipids may surround the nucleocapsid, depending on the virus concerned. These are usually derived from the host cell, but they may also contain viral proteins which have been coded by viral genes.

The complete structure is known as a **virion** and this is the form that the virus takes when it is outside the host cell. The size of a virion can vary from 10 nm to 400 nm. As a result, most viruses are too small to be seen by a light microscope and require the use of an electron microscope.

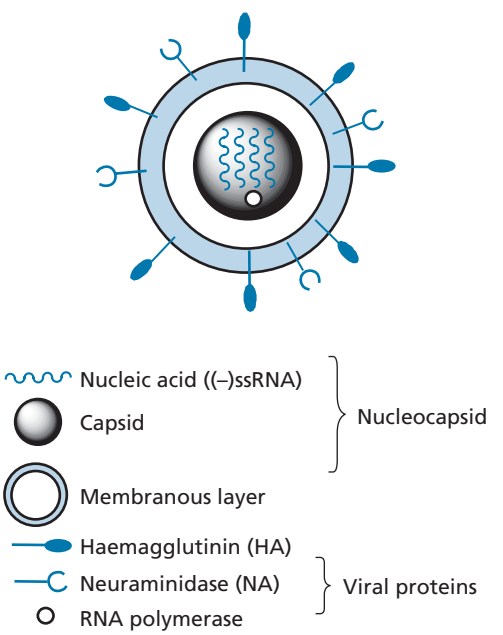


FIGURE 20.1 Diagrammatic representation of the flu virus.

20.3 Life cycle of viruses

The various stages involved in the life cycle of a virus are as follows (Fig. 20.2).

- **Adsorption**: a virion has to first bind to the outer surface of a host cell. This involves a specific molecule on the outer surface of the virion binding to a specific protein or carbohydrate present in the host cell membrane. The relevant molecule on the host cell can thus be viewed as a 'receptor' for the virion. Of course, the host cell has not produced this molecule to be a viral receptor. The molecules concerned are usually glycoproteins which have crucial cellular functions, such as the binding of hormones. However, the virion takes advantage of these, and once it is bound, the next stage can take place—introduction of the viral nucleic acid into the host cell.
- **Penetration and uncoating**: different viruses introduce their nucleic acid into the host cell by different methods. Some inject their nucleic acid through the cell membrane; others enter the cell intact and are then uncoated. This can also happen in a variety of ways. The viral envelope of some virions fuses with the plasma membrane and the nucleocapsid is then introduced into the cell (Fig. 20.2). Other virions are taken into the cell by endocytosis where the cell membrane wraps itself round the virion and is then pinched off to produce a vesicle called an **endosome** (see for example Fig. 20.40). These vesicles then fuse with **lysosomes**, and host cell enzymes aid the virus in the uncoating process. Low endosomal pH also triggers uncoating. In some cases, the viral envelope fuses with the lysosome membrane and the nucleocapsid is released into the cell. Whatever the process, the end result is the release of viral nucleic acid into the cell.
- **Replication and transcription**: viral genes can be defined as *early* or *late*. Early genes take over the host cell such that viral DNA and/or RNA is synthesized. The mechanism involved varies from virus to virus. For example, viruses containing negative ssRNA use a viral enzyme called RNA-dependent RNA polymerase (or transcriptase) to synthesize mRNA which then codes for viral proteins.
- **Synthesis and assembly of nucleocapsids**: late genes direct the synthesis of capsid proteins and these self-assemble to form the capsid. Viral nucleic acid is then taken into the capsid to form the nucleocapsid.

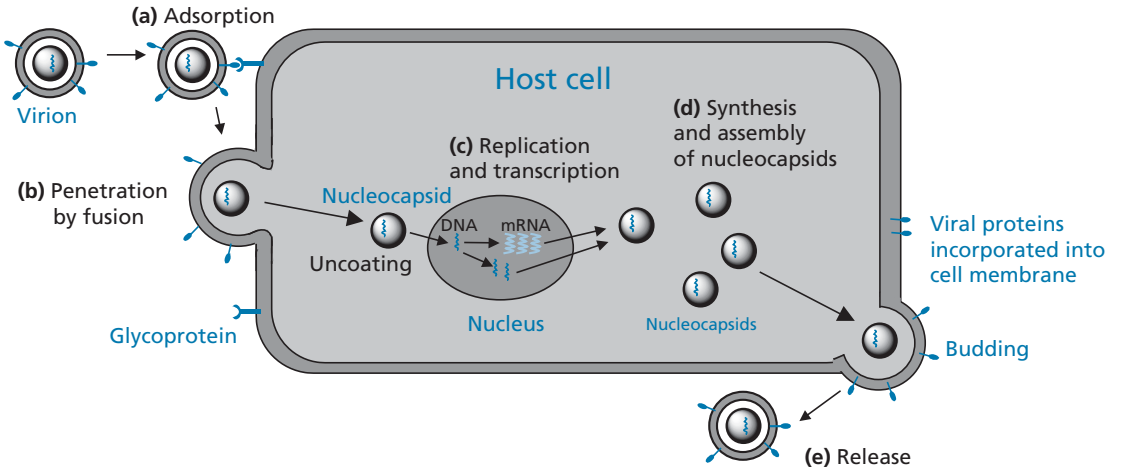


FIGURE 20.2 Life cycle of a DNA virus such as herpes simplex.

- **Virion release**: naked virions (those with no outer layers round the nucleocapsid) are released by cell lysis where the cell is destroyed. In contrast, viruses with envelopes are usually released by a process known as **budding** (Fig. 20.2). Viral proteins are first incorporated into the host cell's plasma membrane. The nucleocapsid then binds to the inner surface of the cell membrane and, at the same time, viral proteins collect at the site and host cell proteins are excluded. The plasma membrane containing the viral proteins then wraps itself round the nucleocapsid and is pinched off from the cell to release the mature virion.

The life cycle stages of herpes simplex, HIV, and flu virus are illustrated in Figs. 20.2, 20.11, and 20.40 respectively.

20.4 Vaccination

Vaccination is the preferred method of protection against viral disease and has proved extremely successful against childhood diseases such as polio, measles, and mumps, as well as historically serious diseases such as smallpox and yellow fever. The first successful vaccination was carried out by Edward Jenner in the eighteenth century. He observed that a milkmaid, who had contracted the less virulent cowpox, was immune to smallpox. Therefore, Jenner inoculated people with material from cowpox lesions and discovered that they, too, gained immunity from smallpox. Since then, many other vaccines have been developed. Perhaps the most controversial vaccination in recent years has been the **MMR vaccine**—a combination of three separate vaccinations administered to young children to provide protection against measles, mumps, and rubella. Unfortunately, deep concerns were raised in the UK as a result of a publication in the *Lancet* that linked the vaccine to an increased risk of autism. This article was eventually discredited as fraudulent.

Vaccination works by introducing the body to foreign material which bears molecular similarity to some component of the virus, but which lacks its infectious nature or toxic effects. The body then has the opportunity to recognize the molecular fingerprint of the virus (i.e. specific **antigens**) and the immune system is primed to attack the virus should it infect the body. Usually a killed or weakened version of the virus is administered so that it does not lead to infection itself. Alternatively, fragments of the virus (subunit vaccines) can be used if they display a characteristic antigen. Vaccination is a preventive approach and is not usually effective on patients who have already become infected.

Vaccines are currently under investigation for the prevention or treatment of HIV, dengue fever, genital herpes, and haemorrhagic fever caused by the Ebola virus. However, there are difficulties surrounding the HIV and flu viruses, because rapid gene mutation in these viruses results in constant changes to the amino acid composition of glycoproteins normally present on the viral surface. Because these glycoproteins are the important antigens that trigger the immune response, any changes in their structure 'disguise' the virus, and the body's primed immune system fails to recognize it.

Another problem concerning vaccination relates to patients with a weakened immune response. The main categories of patients in this situation are cancer patients undergoing chemotherapy, patients undergoing organ transplants (where the immune system has been deliberately suppressed to prevent organ rejection), and AIDS patients. Vaccination in these patients is less likely to be effective because of the weakened immune response.

20.5 Antiviral drugs: general principles

Antiviral drugs are useful in tackling viral diseases where there is a lack of an effective vaccine or where infection has already taken place. The life cycle of a virus means that for most of its time in the body, it is within a host cell and is effectively disguised both from the immune system and from circulating drugs. As it also uses the host cell's own biochemical mechanisms to multiply, the number of potential drug targets that are unique to the virus is more limited than those for invading microorganisms. Thus, the search for effective antiviral drugs has proved more challenging than that for antibacterial drugs. Indeed, the first antiviral agents appeared relatively late on in the 1960s and only three clinically useful antiviral drugs were in use during the early 1980s. Early antiviral drugs included **idoxuridine** and **vidarabine** for herpes infections, and **amantadine** for influenza A.

Since then, progress has accelerated for two principal reasons: (i) the need to tackle the AIDS pandemic and (ii) the increased understanding of viral infectious mechanisms resulting from viral genomic research.

In 1981, it was noticed that homosexual men were unusually susceptible to diseases such as pneumonia and fungal infections—ailments which were previously only associated with patients whose immune response had been weakened. The problem soon reached epidemic proportions and it was discovered that a virus (the **human immunodeficiency virus**—**HIV**) was responsible. It was found that this virus infected **T-cells**—cells which are crucial to the immune response—and was therefore directly attacking the immune system. With a weakened immune system, infected patients proved susceptible to a whole range of opportunistic secondary diseases resulting in the term **acquired immune deficiency syndrome** (**AIDS**). This discovery led to a major research effort into understanding the disease and counteracting it—an effort which kick-started more general research into antiviral chemotherapy. Fortunately, the tools needed to carry out effective research appeared on the scene at about the same time with the advent of viral genomics. The full genome of any virus can now be determined quickly and compared with those of other viruses, allowing the identification of how the genetic sequence is split into genes. This, in turn, helps to identify viral proteins as potential drug targets. Standard genetic engineering methods permit the production of pure copies of the target protein by inserting the viral gene into a bacterial cell thus providing sufficient quantities of the protein to be isolated, and studied (section 6.4).

Good drug targets are proteins which are likely to have the following characteristics:

- they are important to the life cycle of the virus such that their inhibition or disruption has a major effect on infection;
- they bear little resemblance to human proteins, thus increasing the chances of good selectivity and minimal side effects;
- they are common to a variety of different viruses and have a specific region which is identical in its amino acid composition. This makes the chances of developing a drug with broad antiviral activity more likely;
- they are important to the early stages of the viral life cycle, which means that their disruption/inhibition reduces the chances of symptoms and of the virus spreading through the body.

Most antiviral drugs in use today act against HIV, herpesviruses (responsible for a variety of ailments, including cold sores and encephalitis), hepatitis B, and hepatitis C. Diseases such as herpes and HIV are chronic in developed countries, and intensive research has been carried out to develop drugs to combat them. In contrast, less research is carried out on viral diseases that are prevalent in developing countries; for example tropical (dengue) and haemorrhagic (Ebola) fevers.

Most antiviral drugs in use today disrupt critical stages of the virus life cycle or the synthesis of virus-specific nucleic acids. Excluding drugs developed for the treatment of HIV, more drugs are available for the treatment of DNA viruses than for RNA viruses. Few drugs show a broad activity against both DNA and RNA viruses.

Studies of the human genome are also likely to be useful for future research. The identification of human proteins which stimulate the body's immune response or the production of antibodies would provide useful leads for the development of drugs that would have an antiviral effect by acting as immunomodulators.

KEY POINTS

- Viruses pose a serious health threat and there is a need for new antiviral agents.
- Viruses consist of a protein coat surrounding nucleic acid, which is either RNA or DNA. Some viruses have an outer membranous coat that is derived from the host cell.
- Viruses are unable to self-multiply and need to enter a host cell in order to do so.
- Vaccination is effective against many viruses, but is less effective against viruses which readily mutate.
- Research into antiviral drugs has increased in recent years as a result of the AIDS epidemic and the need to find drugs to combat it.
- Antiviral research has been aided by advances in viral genomics and genetic engineering, as well as by the use of X-ray crystallography and molecular modelling.

20.6 Antiviral drugs used against DNA viruses

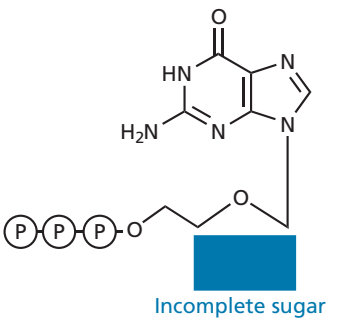
Most of the drugs which are active against DNA viruses have been developed against herpesviruses to combat diseases such as cold sores, genital herpes, chicken pox, shingles, eye diseases, mononucleosis, Burkitt's lymphoma, and Kaposi's sarcoma. Nucleoside analogues have been particularly effective.

20.6.1 Inhibitors of viral DNA polymerase

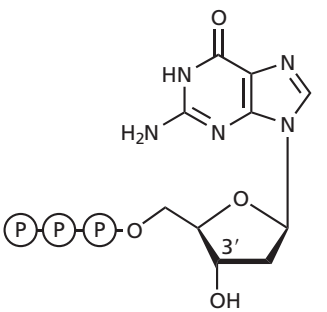
Aciclovir was discovered by compound screening and was introduced into the market in 1981. Aciclovir has a nucleoside-like structure and contains the same nucleic acid base as deoxyguanosine. However, it lacks the complete sugar ring. In virally infected cells, it is phosphorylated in three stages to form a triphosphate which is the active agent, and so aciclovir is a prodrug (Fig. 20.3).

Nucleotide triphosphates are the building blocks for DNA replication where a new DNA strand is constructed using a DNA template—a process catalysed by the enzyme **DNA polymerase**. Aciclovir triphosphate prevents DNA replication in two ways. Firstly, it is sufficiently similar to the normal deoxyguanosine triphosphate building block (Fig. 20.4) that it can bind to DNA polymerase and inhibit it. Secondly, DNA polymerase can catalyse the attachment of the aciclovir nucleotide to the growing DNA chain. As the sugar unit is incomplete and lacks the required hydroxyl group normally present at position 3′ of the sugar ring, the nucleic acid chain cannot be extended any further. Thus, the drug acts as a **chain terminator** (see section 9.5).

However, what is to stop aciclovir triphosphate inhibiting DNA polymerase in normal, uninfected cells? The answer lies in the fact that aciclovir is only converted to



Aciclovir triphosphate



Deoxyguanosine triphosphate

FIGURE 20.4 Comparison of aciclovir triphosphate and deoxyguanosine triphosphate.

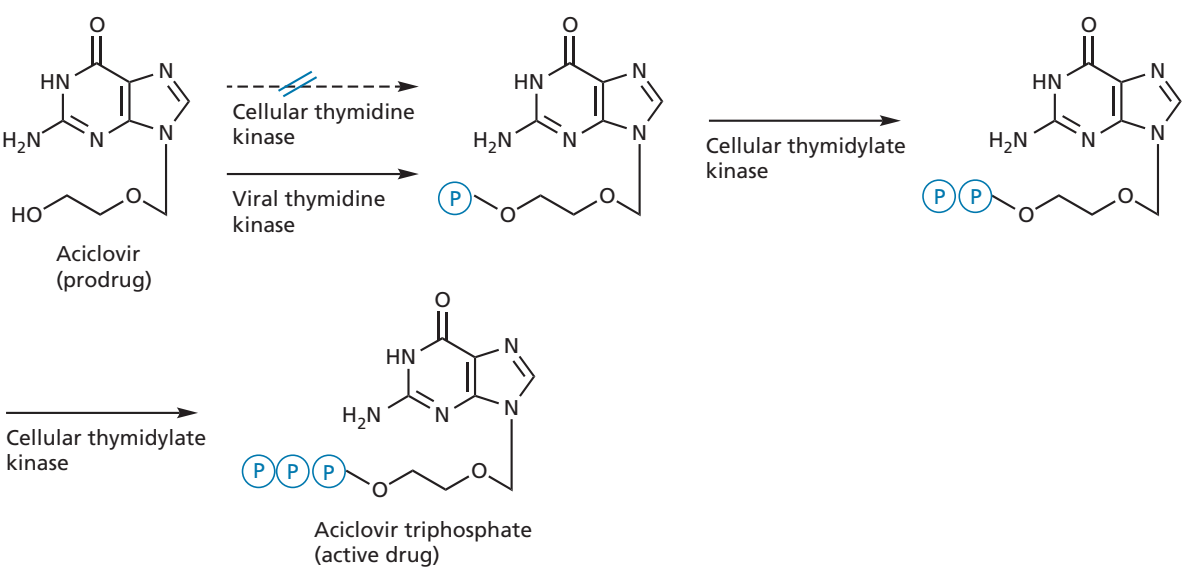


FIGURE 20.3 Activation of aciclovir. (P) represents phosphate groups.

the active triphosphate in infected cells. The explanation for this lies in the first phosphorylation reaction catalysed by the enzyme thymidine kinase (Fig. 20.3). Although this enzyme is present in host cells, the herpes virus carries its own version. It turns out that viral thymidine kinase is 100 times more effective at converting aciclovir to its monophosphate than host cell thymidine kinase. Once formed, the monophosphate is converted to the active triphosphate by cellular enzymes. Therefore, in normal, uninfected cells, aciclovir is a poor substrate for cellular thymidine kinase and remains as the prodrug. This, along with the fact that there is a selective uptake of aciclovir by infected cells, explains its excellent activity and much reduced toxicity relative to previous drugs. Another feature which enhances its safety is that aciclovir triphosphate shows a 50-fold selective action against viral DNA polymerases relative to cellular polymerases.

The oral bioavailability of aciclovir is quite low (15–30%). To overcome this, various prodrugs were developed to increase water solubility. **Valaciclovir** (Fig. 20.5) is an L-valyl ester prodrug absorbed from the gut far more effectively than aciclovir. However, the prodrug has similar polarity and ionization to aciclovir, and so the prodrug is no more able to cross the cell membranes of the gut wall by passive diffusion than aciclovir. Moreover, poorer absorption is observed if D-valine is used for the prodrug instead of L-valine, suggesting that a specific binding interaction is involved in the absorption process. This implies that the prodrug is actively transported by transport proteins in the gut, and that the valine allows the prodrug to be recognized and bound by these proteins. Transport proteins normally responsible for transporting dipeptides across the cell wall have been implicated in this process, i.e. the **human intestinal proton-dependent oligopeptide transporter-1** (**hPEPT-1**) and **human intestinal di-/tripeptide transporter-1** (**HPT-1**). Once valaciclovir is absorbed, it is hydrolysed to aciclovir in the liver and gut wall. **Desciclovir** (Fig. 20.5) is a prodrug of aciclovir which lacks the carbonyl group at position 6 of the purine ring and is more water soluble. Once in the blood supply, metabolism by cellular **xanthine oxidase** oxidizes the 6-position to give aciclovir.

Ganciclovir (Fig. 20.5) is an analogue of aciclovir and bears an extra hydroxymethylene group; **valganciclovir** acts as a prodrug for this compound. **Penciclovir** and its prodrug **famciclovir** (Fig. 20.6) are analogues of ganciclovir. In famciclovir, the two alcohol groups of penciclovir are masked as esters making the structure less polar, resulting in better absorption. Once absorbed, the acetyl groups are hydrolysed by esterases and the purine ring is oxidized by **aldehyde oxidase** in the liver to generate penciclovir. Phosphorylation reactions then take place in virally infected cells, as described previously.

Some viruses are immune from the action of these antiviral agents because they lack the enzyme thymidine kinase. As a result, phosphorylation fails to take place. **Cidofovir** was designed to combat this problem (Fig. 20.7). It is an analogue of **deoxycytidine 5-monophosphate** where the sugar and phosphate groups have been replaced by an acyclic group and a phosphonomethylene group respectively. The latter group acts as a bioisostere for the phosphate group and is used because the phosphate group itself would be susceptible to enzymatic hydrolysis. As a phosphate equivalent is now present, the drug does not require thymidine kinase to become activated. Two more phosphorylations can now take place catalysed by cellular kinases, to convert cidofovir to the active 'triphosphate'.

In contrast to aciclovir, **idoxuridine**, **trifluridine**, and **vidarabine** (Fig. 20.8) are phosphorylated equally well by viral and cellular thymidine kinases, and so there is less selectivity for virally infected cells. As a result, these drugs have more toxic side effects. Idoxuridine, like trifluridine, is an analogue of deoxythymidine and was the first nucleoside-based antiviral agent licensed in the USA. The triphosphate inhibits viral DNA polymerase, as well as thymidylate synthetase. **Vidarabine** (Fig. 20.8) contains an arabinoside sugar ring and was developed

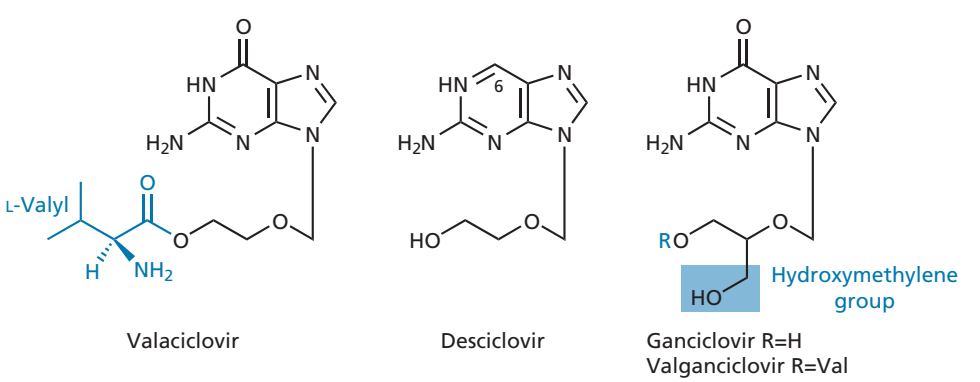


FIGURE 20.5 Prodrugs and analogues of aciclovir.

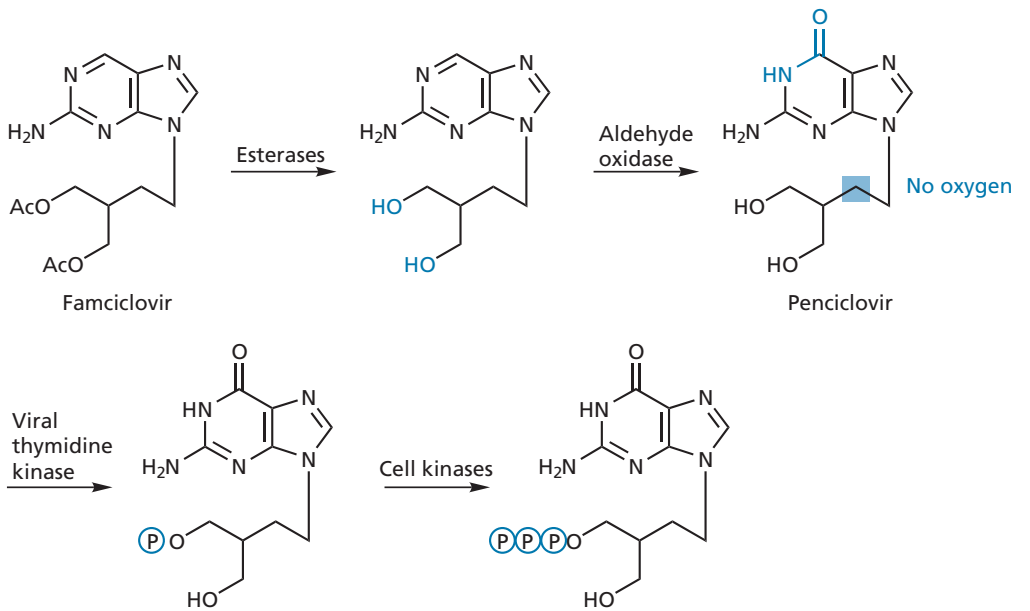


FIGURE 20.6 Penciclovir and famciclovir. (P) represents a phosphate group.

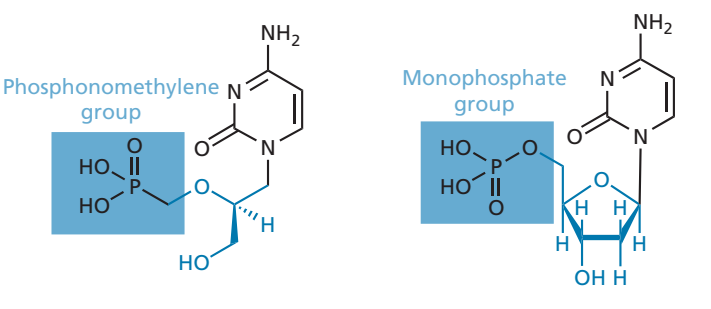


FIGURE 20.7 Comparison of cidofovir and deoxycytidine monophosphate.

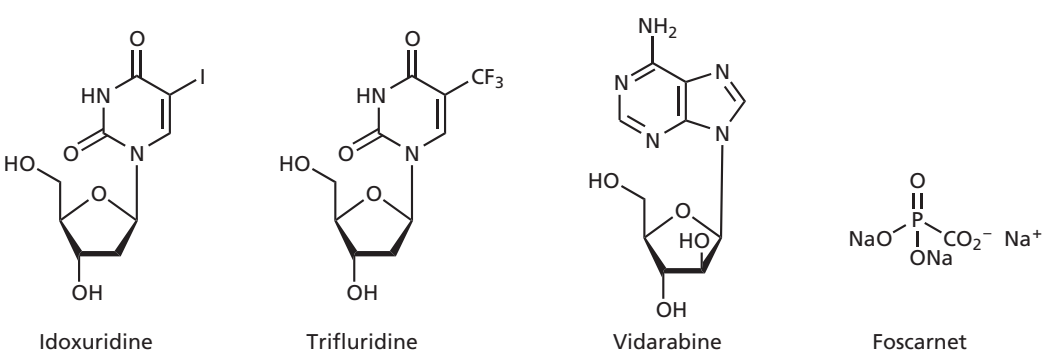


FIGURE 20.8 Miscellaneous antiviral agents.

from a natural product isolated from a marine sponge. **Foscarnet** (Fig. 20.8) was discovered in the 1960s and inhibits viral DNA polymerase. However, it is non-selective and toxic. It also has difficulty crossing cell membranes due to its high charge.

20.6.2 Inhibitors of tubulin polymerization

The plant product **podophyllotoxin** (Fig. 20.9) has been used clinically to treat genital warts caused by the DNA virus **papillomavirus**, but it is not as effective as

BOX 20.1 Clinical aspects of viral DNA polymerase inhibitors

Aciclovir represented a revolution in the treatment of herpes infections, being the first relatively safe, non-toxic drug to be used systemically. It is used for the treatment of infections due to herpes simplex 1 (HSV1) and 2 (HSV2) (i.e. herpes simplex encephalitis and genital herpes), as well as **varicella-zoster viruses** (**VZV**) (i.e. chickenpox and shingles). Unfortunately, strains of herpes are appearing which are resistant to aciclovir. This can arise as a result of mutations, either of the viral thymidine kinase enzyme, such that it no longer phosphorylates aciclovir, or of viral DNA polymerase, such that it no longer recognizes the activated drug. Aciclovir is not effective against all types of herpes virus. There are eight herpes viruses that are divided into three subfamilies. Aciclovir is effective against the α-subfamily but not the β-subfamily because the latter produces a different thymidine kinase that fails to phosphorylate the drug.

Valaciclovir is a valine prodrug of aciclovir and is particularly useful in the treatment of VZV infections. When this prodrug is given orally, blood levels of aciclovir are obtained which are equivalent to those obtained by intravenous administration.

Desciclovir is another prodrug for aciclovir but is somewhat more toxic, thus limiting its potential.

Ganciclovir is phosphorylated by thymidine kinases produced by both the α- and β-subfamilies of herpesvirus, and can be used against both viruses. Unfortunately, the drug is not as safe as aciclovir because it can be incorporated into cellular DNA. Nevertheless, it can be used for the treatment of **cytomegalovirus** (**CMV**) infections. This is a virus which causes eye infections and can lead to blindness. Aciclovir is not effective in this infection because CMV does not encode a viral thymidine kinase. Ganciclovir, however, can be converted to its monophosphate by kinases other than thymidine kinase. As ganciclovir has a low oral bioavailability, the valine prodrug **valganciclovir** has been introduced for the treatment of CMV infections.

Penciclovir has essentially the same spectrum of activity as aciclovir, but has better potency, faster onset, and longer duration of action. It is used topically for the treatment of cold sores (HSV-1), and intravenously for the treatment of HSV in immunocompromised patients. Like aciclovir, penciclovir has poor oral bioavailability and is poorly absorbed from the gut because of its polarity. Therefore, **famciclovir** is used as a prodrug for better absorption.

Cidofovir is a broad-spectrum antiviral agent which shows selectivity for viral DNA polymerase and is used to treat retinal inflammation caused by CMV. Unfortunately, the drug is extremely polar and has a poor oral bioavailability (5%). It is also toxic to the kidneys, but this can be reduced by co-administering **probenecid** (section 19.5.1.9).

Idoxuridine can be used for the topical treatment of herpes keratitis, but **trifluridine** is the drug of choice for this disease because it is effective at lower dose frequencies. **Vidarabine** was an early antiviral drug, but aciclovir is now generally preferred.

Foscarnet is used in the treatment of CMV retinitis where it is approximately equal in activity to ganciclovir. It can also be used in immunocompromised patients for the treatment of HSV and VZV strains which prove resistant to aciclovir. It does not undergo metabolic activation.

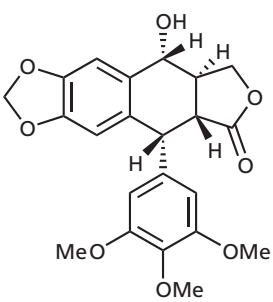


FIGURE 20.9 Podophyllotoxin.

imiquimod (section 20.11.4). It is a powerful inhibitor of tubulin polymerization (sections 2.7.1 and 10.2.2).

20.6.3 Antisense therapy

Fomivirsen (Fig. 20.10) is the first, and so far the only, DNA antisense molecule that has been approved as an

d(*P*-thio)(G-C-G-T-T-T-G-C-T-C-T-T-C-T-T-C-T-T-G-C-G)

FIGURE 20.10 Fomivirsen.

antiviral agent. It consists of 21 nucleotides with a phosphonothioate backbone rather than a phosphate backbone to increase the metabolic stability of the molecule (section 14.10). The drug blocks the translation of viral RNA and is used against retinal inflammation caused by CMV in AIDS patients. Because of its high polarity, it is administered as an ocular injection (**intravitreal injection**).

KEY POINTS

- Nucleoside analogues have been effective antiviral agents against DNA viruses, mainly herpesviruses.
- Nucleoside analogues are prodrugs, which are activated by phosphorylation to a triphosphate structure. They have a dual mechanism of action as viral DNA polymerase inhibitors and DNA chain terminators.

- Nucleoside analogues show selectivity for virally infected cells over normal cells when viral thymidine kinase is required to catalyse the first of three phosphorylation steps. They are taken up more effectively into virally infected cells and their triphosphates show selective inhibition for viral DNA polymerases over cellular DNA polymerases.
- Agents containing a bioisostere for a phosphate group can be used against DNA viruses lacking thymidine kinase.
- Inhibitors of tubulin polymerization have been used against DNA viruses.
- An antisense molecule has been designed as an antiviral agent.

20.7 Antiviral drugs acting against RNA viruses: HIV

20.7.1 Structure and life cycle of HIV

HIV (Fig. 20.11) is an example of a group of viruses known as the **retroviruses**. There are two variants of HIV: HIV-1 is responsible for AIDS in the USA, Europe, and Asia, whereas HIV-2 occurs mainly in western Africa. HIV has been studied extensively over the last 20 years and a vast research effort has resulted in a variety of antiviral drugs which have proved successful in slowing down the disease, but not eradicating it. At present, most clinically useful antiviral drugs act against two targets: the viral enzymes **reverse transcriptase** and **protease**. There is a need to develop effective drugs against a third target and a good knowledge of the life cycle of HIV is essential in identifying suitable targets (Fig. 20.11).

HIV is an RNA virus which contains two identical strands of (+) ssRNA within its capsid. Also present are the viral enzymes reverse transcriptase and **integrase**, as well as other proteins called **p6** and **p7**. The capsid is made up of protein units known as **p24**; surrounding the capsid there is a layer of matrix protein (**p17**), then a membranous envelope which originates from host cells and which contains the viral glycoproteins **gp120** and **gp41**. Both of these proteins are crucial to the processes of adsorption and penetration. Gp41 traverses the envelope and is bound non-covalently to gp120, which projects from the surface. When the virus approaches

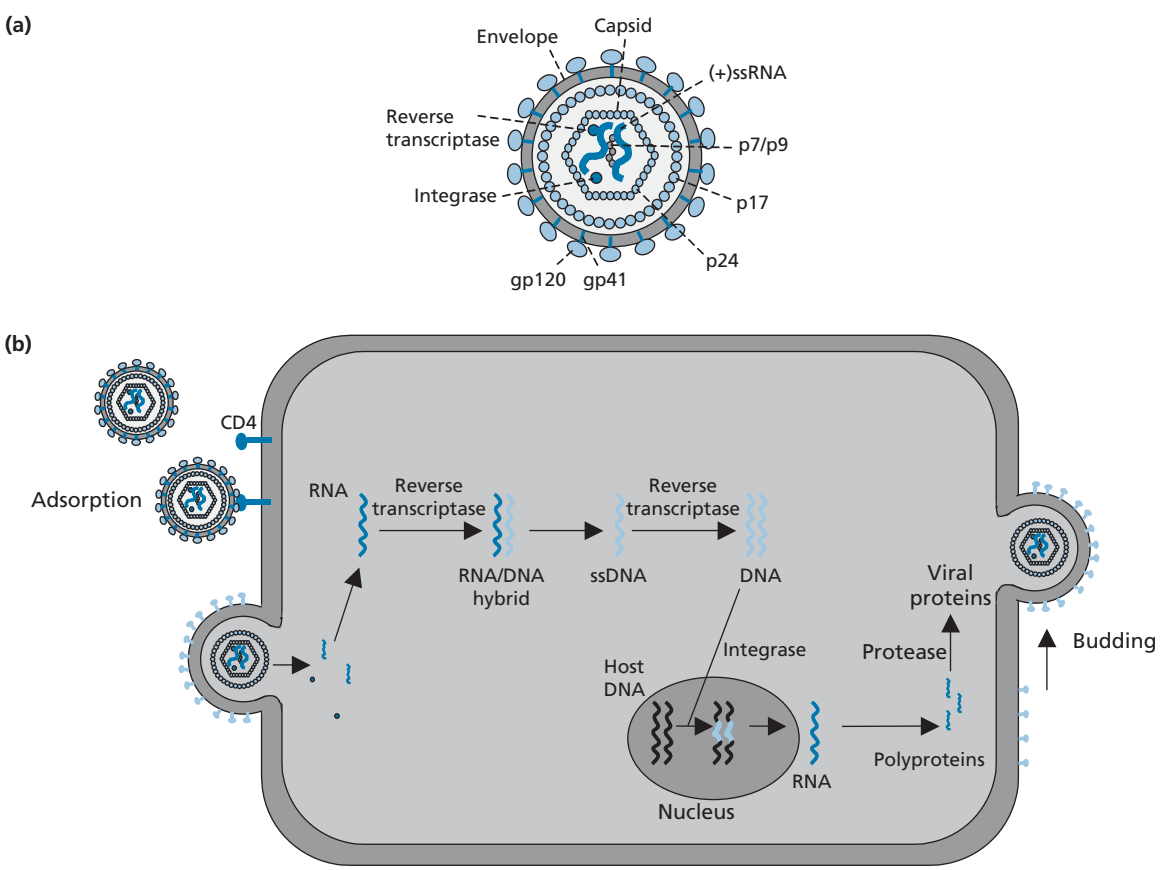


FIGURE 20.11 (**a**)Structure of HIV particle (p = protein; gp = glycoprotein) and (**b**) life cycle of HIV in a host T-cell.

the host cell, gp120 interacts and binds with a transmembrane protein called **CD4**, which is present on host T-cells. The gp120 proteins then undergo a conformational change which allows them to bind simultaneously to **chemokine receptors** (**CCR5** and **CXCR4**) on the host cell (not shown). Further conformational changes peel away the gp120 protein allowing the viral protein gp41 to reach the surface of the host cell and anchor the virus to the surface. The gp41 then undergoes a conformational change and pulls the virus and the cell together so that their membranes can fuse.

Once fusion has taken place, the HIV nucleocapsid enters the cell. Disintegration of the protein capsid then takes place, probably aided by the action of a viral enzyme called protease. Viral RNA and viral enzymes are then released into the cell cytoplasm. The released viral RNA is not capable of coding directly for viral proteins or of self-replication. Instead, it is converted into DNA and incorporated into the host cell DNA. The conversion of RNA into DNA is not a process that occurs in human cells, so there are no host enzymes to catalyse the process. Therefore, HIV carries its own enzyme—**reverse transcriptase**—to do this. This enzyme is a member of a family of enzymes known as the DNA polymerases, but is unusual in that it can use a RNA strand as a template. The enzyme first catalyses the synthesis of a DNA strand using viral RNA as a template. This leads to a (+)RNA–(−) DNA hybrid. Reverse transcriptase catalyses the degradation of the RNA strand then uses the remaining DNA strand as a template to catalyse the synthesis of dsDNA (**proviral DNA**). Proviral DNA is now spliced into the host cell's DNA—a process catalysed by the viral protein **integrase**. Once the proviral DNA has been incorporated into host DNA, it is called the **provirus** and can remain dormant in host cell DNA until activated by cellular processes. When that occurs, transcription of the viral genes *env, gag*, and *pol* takes place to produce viral RNA, some of which will be incorporated into new virions, and the rest of which is used in translation to produce three large, non-functional polyproteins, one derived from the *env* gene, one from the *gag* gene, and the other from the *gag–pol* genes. The first of these polyproteins is cleaved by cellular proteinases and produces the viral glycoproteins (gp120 and gp41), which are incorporated into the cell membrane. The remaining two polypeptides (**Pr55** and **Pr160**) remain intact and move to the inner membrane surface. The viral glycoproteins in the cell membrane also concentrate in this area and cellular proteins are excluded. Budding then takes place to produce an immature, membrane-bound virus particle. During the budding process a viral enzyme called protease is released from the gag–pol polypeptide. This is achieved by the protease enzyme autocatalysing the cleavage of susceptible peptide bonds linking it to the rest of the polypeptide. Once released, the protease enzyme dimerizes and cleaves the remaining polypeptide chains to release reverse transcriptase, integrase, and viral structural proteins. The capsid proteins now self-assemble to form new nucleocapsids containing viral RNA, reverse transcriptase, and integrase.

It has been observed that a viral protein called **Vpu** has an important part to play in the budding process. Vpu binds to the host membrane protein CD4 and triggers a host enzyme to tag the CD4 protein with a protein called **ubiquitin**. Proteins that are tagged with ubiquitin are marked out for destruction by the host cell and so the CD4 proteins in the host cell are removed. This is important as the CD4 proteins could complex with the newly synthesized viral proteins gp120 and prevent the assembly of the new viruses.

20.7.2 Antiviral therapy against HIV

Until 1987, no anti-HIV drug was available, but an understanding of the life cycle of HIV has led to the identification of several possible drug targets. At present, most drugs that have been developed act against the viral enzymes reverse transcriptase and protease. However, a serious problem with the treatment of HIV is the fact that the virus undergoes mutation extremely easily. This results in rapid resistance to antiviral drugs. Experience has shown that treatment of HIV with a single drug has a short-term benefit, but, in the long term, the drug serves only to select mutated viruses which are resistant. As a result, current therapy involves combinations of different drugs acting on both reverse transcriptase and protease (Box 20.2). This has been successful in delaying the progression to AIDS and increasing survival rates, but there is a need to develop effective drugs against a third target.

The demands on any HIV drug are immense, especially as it is likely to be taken over long periods of time. It must have a high affinity for its target (in the picomolar range) and be effective in preventing the virus multiplying and spreading. It should show low activity for any similar host targets in the cell, and be safe and well tolerated. It must be active against as large a variety of viral isolates as possible or else it only serves to select resistant variants. It needs to be synergistic with other drugs used to fight the disease and be compatible with other drugs used to treat opportunistic diseases and infections arising from the weakened immune response. The drug must stay above therapeutic levels within the infected cell and in the circulation. It must be capable of being taken orally and with a minimum frequency of doses, and it should preferably be able to cross the blood–brain barrier in case the virus lurks in the brain. Finally, it must be inexpensive as it is likely to be used for the lifetime of the patient.

BOX 20.2 Clinical aspects of antiviral drugs used against HIV

There is no cure for HIV infection, but anti-HIV drugs can halt, or slow, the rate at which the disease develops, leading to a significant increase in life expectancy. Unfortunately, the drugs used have toxic side effects, which is particularly significant as patients have to take these drugs for the rest of their lives. This means that patients have to be monitored constantly. A combination of drugs acting against two different enzyme targets is used—**highly active antiretroviral therapy** (**HAART**). When choosing which drugs to use, it is important to ensure that they have a synergistic or additive effect, and that they are compatible in terms of their toxic properties.

Currently, **protease inhibitors** (PIs) (section 20.7.4) are used with **reverse transcriptase inhibitors** (section 20.7.3) (divergent therapy) or with another PI (convergent therapy). A combination of two **nucleoside reverse transcriptase inhibitors** (NRTIs) plus a PI is recommended, but one can also use two PIs with an NRTI, or a **non-nucleoside reverse transcriptase inhibitor** (**NNRTI**) with two NRTIs. For example, the NNRTI efavirenz is used along with the NRTIs emtricitabine and tenofovir.

The NRTIs which can be used against HIV are zidovudine, didanosine, zalcitabine, stavudine, lamivudine, emtricitabine, tenofovir disoproxil, and abacavir. NNRTIs used against HIV are nevirapine, delavirdine, and efavirenz. The PIs which are used are saquinavir, ritonavir, indinavir, nelfinavir, amprenavir, atazanavir, darunavir, fosamprenavir, lopinavir, and tipranavir.

The fusion inhibitor enfuvirtide has now been approved as an anti-HIV drug, and acts against a different target from the conventional anti-HIV drugs. It can be included alongside conventional drugs if the disease fails to respond to standard HAART therapy. The first integrase inhibitor has also been approved for clinical use.

As a result of HAART, the death rate from AIDS-related deaths appears to be slowing. For example, 1.8 million deaths were recorded in 2009 compared with 2.4 million in 2004.

20.7.3 Inhibitors of viral reverse transcriptase

20.7.3.1 Nucleoside reverse transcriptase inhibitors

As the enzyme reverse transcriptase is unique to HIV, it serves as an ideal drug target. Nevertheless, the enzyme is still a DNA polymerase and care has to be taken that inhibitors do not have a significant inhibitory effect on cellular DNA polymerases. Various nucleoside-like structures have proved useful as antiviral agents. The vast majority of these are not active themselves, but are phosphorylated by three cellular enzymes to form an active nucleotide triphosphate. This is the same process described previously in section 20.6.1, but there is one important difference: cellular enzymes are required to catalyse all three phosphorylations because HIV does not produce a viral kinase.

Zidovudine (Fig. 20.12) was developed originally as an anticancer agent but was the first drug to be approved for use in the treatment of AIDS. It is an analogue of deoxythymidine where the sugar 3′-hydroxyl group has been replaced by an azido group. On conversion

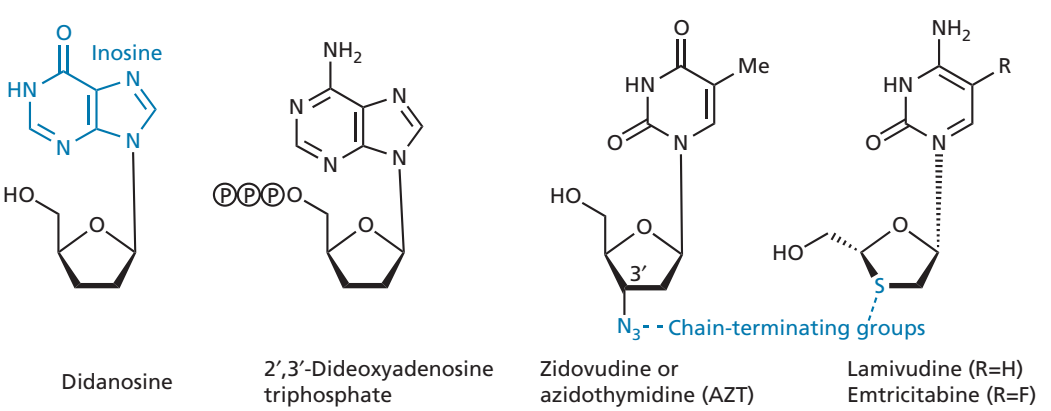


FIGURE 20.12 Inhibitors of viral reverse transcriptase. Ⓟ = phosphate.

to the triphosphate, it inhibits reverse transcriptase. Furthermore, the triphosphate is attached to the growing DNA chain. Since the sugar unit has an azide substituent at the 3′ position of the sugar ring, the nucleic acid chain cannot be extended any further.

Didanosine (Fig. 20.12) was the second anti-HIV drug approved for use in the USA (1988). Its activity was unexpected as the nucleic acid base present is inosine—a base which is not incorporated naturally into DNA. However, a series of enzyme reactions converts this compound into 2′,3′-dideoxyadenosine triphosphate which is the active drug. Studies of the target enzyme's active site led to the development of **lamivudine** and **emtricitabine** (Fig. 20.12) (analogues of deoxycytidine where the 3′ carbon has been replaced by sulphur). Other clinically useful NRTIs used against HIV and/or hepatitis B include **abacavir** (the only guanosine analogue), **stavudine**, and **zalcitabine** (Fig. 20.13). **Tenofovir disoproxil** and **adefovir dipivoxil** are prodrugs of modified nucleosides. Both structures contain a monophosphate group protected by two extended esters. Hydrolysis *in vivo* reveals the phosphate group which can then be phosphorylated to the triphosphate as described previously.

20.7.3.2 Non-nucleoside reverse transcriptase inhibitors (NNRTIs)

The NNRTIs (Fig. 20.14) are generally hydrophobic molecules that bind to an allosteric binding site which is hydrophobic in nature. Since the allosteric binding site is separate from the substrate binding site, the NNRTIs are non-competitive, reversible inhibitors. They include first-generation NNRTIs, such as **nevirapine** and **delavirdine**, as well as second-generation drugs, such as **efavirenz**, **etravirine**, and **rilpivirine**. X-ray crystallographic studies on inhibitor–enzyme complexes show that the allosteric binding site is adjacent to the substrate binding site. Binding of a NNRTI to the allosteric site results in an induced fit which locks the neighbouring substrate-binding site into an inactive conformation. Unfortunately, rapid resistance emerges as a result of mutations in the NNRTI binding site—the most common being the replacement of Lys-103 with asparagine. This mutation is called K103N and is defined as a **pan-class resistance mutation**. The resistance problem can be countered by combining an NNRTI with an NRTI from the start of treatment. The two types of drugs can be used together as the binding sites are distinct.

Nevirapine was developed from a lead compound discovered through a random screening programme and has a rigid butterfly-like conformation that makes it chiral. One 'wing' interacts through hydrophobic and van der Waals interactions with aromatic residues in the binding site, while the other wing interacts with aliphatic residues. The other NNRTI inhibitors bind to the same pocket and appear to function as π electron donors to aromatic side chain residues.

Delavirdine was developed from a lead compound discovered by a screening programme of 1500 structurally diverse compounds. It is larger than other NNRTIs and extends beyond the normal pocket such that it projects into surrounding solvent. The pyridine region and isopropylamine groups are the most deeply buried parts of the molecule and interact with tyrosine and tryptophan residues. There are also extensive hydrophobic contacts. Unlike other first-generation NNRTIs, there is hydrogen bonding to the main peptide chain next to Lys-103. The indole ring of delavirdine interacts with Pro-236, and mutations involving Pro-236 lead to resistance. Analogues having a pyrrole ring in place of indole may avoid this problem.

Test your understanding and practise your molecular modelling with Exercise 20.1.

Second-generation NNRTIs were developed specifically to find agents that were active against resistant variants,

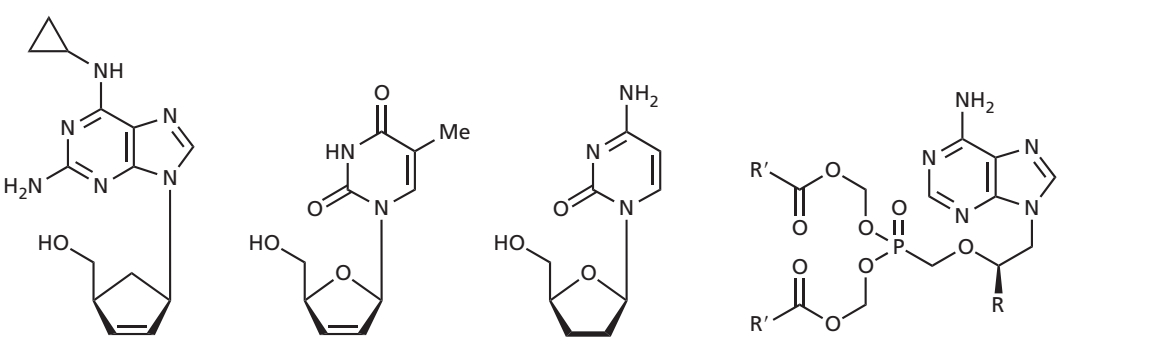


FIGURE 20.13 Further inhibitors of viral reverse transcriptase.

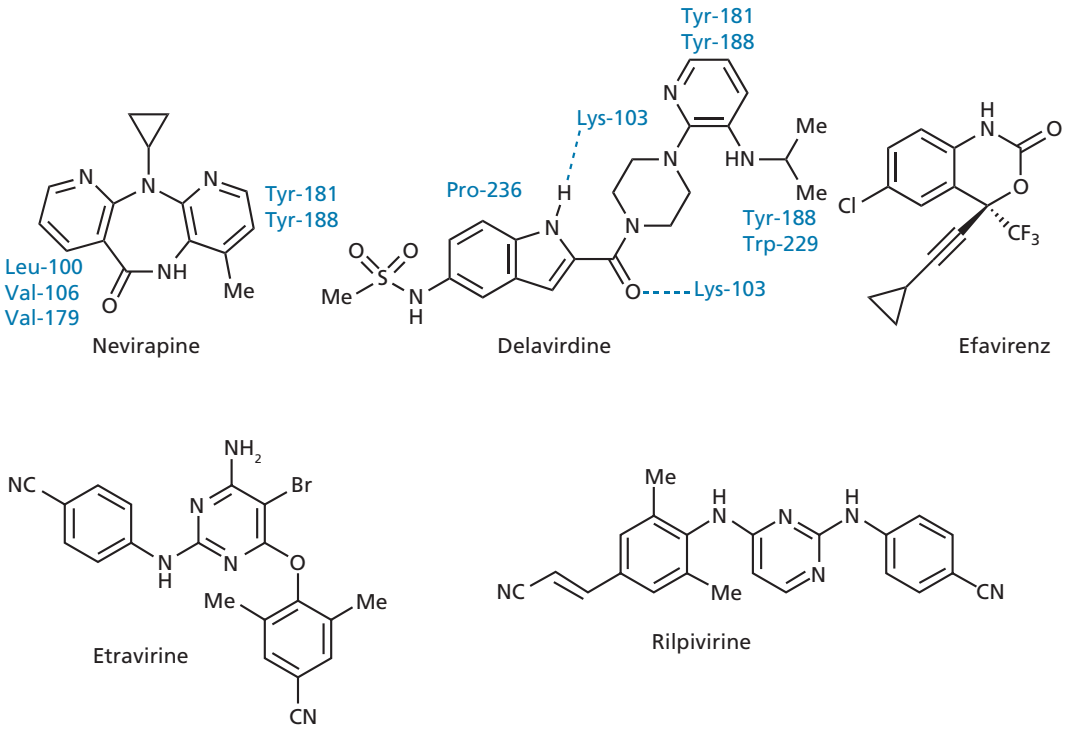


FIGURE 20.14 Non-nucleoside reverse transcriptase inhibitors in clinical use (interactions with amino acids in the binding site are shown in blue).

as well as wild-type viruses. This development has been helped by X-ray crystallographic studies which show how the structures bind to the binding site. It has been shown from sequencing studies that in most of the mutations that cause resistance to first-generation NNRTIs, a large amino acid is replaced by a smaller one, implying that an important binding interaction has been lost. Interestingly, mutations that replace an amino acid with a larger amino acid also appear to be detrimental to the activity of the enzyme, but no mutations have been found which block NNRTIs sterically from entering the binding site.

Efavirenz is a benzoxazinone structure which has activity against many mutated variants but has less activity against the mutated variant K103N. Nevertheless, activity drops less than for nevirapine and a study of X-ray structures of each complex revealed that the cyclopropyl group of efavirenz has fewer interactions with Tyr-181 and Tyr-188 than does nevirapine . Consequently, mutations of these amino acids have a lesser effect on efavirenz than they do on nevirapine. Efavirenz is also a smaller structure and can shift its binding position when K103N mutation occurs, allowing it to form hydrogen bonds to the main peptide chain of the binding site.

X-ray crystallographic studies of enzyme complexes with several second-generation NNRTIs reveal that these agents contain a non-aromatic moiety which interacts with the aromatic residues Tyr-181, Tyr-188, and Trp-229 at the top of the binding pocket. A relatively small bulk and the ability to form hydrogen bonds to the main peptide chain are important as they allow compounds to change their binding mode when mutations occur. The most recent NNRTIs to be approved are etravirine (2008) and rilpivirine (2011).

20.7.4 Protease inhibitors

In the mid 1990s, the use of X-ray crystallography and molecular modelling led to the structure-based design of a series of inhibitors which act on the viral enzyme HIV protease. Like the reverse transcriptase inhibitors, protease inhibitors (PIs) have a short-term benefit when they are used alone, but resistance soon develops. Consequently, combination therapy is now the accepted method of treating HIV infections. When protease and reverse transcriptase inhibitors are used together, the antiviral activity is enhanced and viral resistance is slower to develop.

Unlike the reverse transcriptase inhibitors, the PIs are not prodrugs and do not need to be activated. Therefore, it is possible to use *in vitro* assays involving virally infected cells in order to test their antiviral activity. The protease enzyme can also be isolated, allowing enzyme

BOX 20.3 Clinical aspects of reverse transcriptase inhibitors

Nucleoside reverse transcriptase inhibitors (NRTIs)

NRTIs are currently used as part of the combination therapy for combating HIV. Generally, NRTIs have good oral bioavailability, are bound minimally to plasma proteins, and are excreted through the kidneys. They also act against both HIV-1 and HIV-2. However, they are often associated with toxic side effects. **Zidovudine** was the first anti-HIV drug to reach the market, but can cause severe side effects such as anaemia. **Didanosine** was the second anti-HIV drug approved for use and reached the US market in 1988. However, there is a risk of toxicity to the pancreas. **Abacavir** was approved in 1998 and has been used successfully in children in combination with the protease inhibitors (PIs) nelfinavir and saquinavir. However, life-threatening hypersensitivity reactions have been reported in some patients. **Tenofovir disoproxil** was approved for HIV-1 treatment in 2001. It remains in infected cells longer than many other antiretroviral drugs, allowing for once-daily dosing, but can have toxic effects on the kidneys. It can be used alongside **emtricitabine**, which is relatively free of toxic side effects. Other NRTIs used against HIV include **lamivudine** and **stavudine**. Lamivudine is less toxic than zidovudine and has also been approved for the treatment of hepatitis B.

Zalcitabine is a NRTI which acts against hepatitis B, but long-term toxicity means that it is unacceptable for the treatment of chronic viral diseases which are not life threatening. **Adefovir dipivoxil** was approved by the US Food and Drugs Administration (FDA) in 2002 for the treatment of chronic hepatitis B. It is also active on viruses such as CMV and herpes.

Non-nucleoside reverse transcriptase inhibitors (NNRTIs)

Compared with the NRTIs, the NNRTIs show a higher selectivity for HIV-1 reverse transcriptase over host DNA polymerases. As a result, NNRTIs are less toxic and have fewer side effects. Unfortunately, rapid resistance emerges if a NNRTI is used on its own, but this does not occur if the NNRTI is combined with an NRTI from the start of treatment. NNRTIs are restricted to HIV-1 activity and are generally metabolized by the liver. They can interact with other drugs and bind more strongly to plasma proteins. **Nevirapine**, **efavirenz**, **delavirdine**, **etravirine**, and **rilpivirine** are NNRTIs currently approved by the FDA for the treatment of HIV.

assays to be carried out. In general, the latter are used to measure IC_{50} levels as a measure of how effectively novel drugs inhibit the protease enzyme. The IC_{50} is the concentration of drug required to inhibit the enzyme by 50%. Thus, the lower the IC_{50} value, the more potent the inhibitor. However, a good PI does not necessarily mean a good antiviral drug. In order to be effective, the drug has to cross the cell membrane of infected cells, and so *in vitro* whole-cell assays are often used alongside enzyme studies to check cell absorption. EC_{50} values are a measure of antiviral activity and represent the concentration of compounds required to inhibit 50% of the cytopathic effect of the virus in isolated lymphocytes. Another complication is the requirement for anti-HIV drugs to have a good oral bioavailability (i.e. to be orally active). This is a particular problem with the PIs. As we shall see, most PIs are designed from peptide lead compounds. Peptides are well known to have poor pharmacokinetic properties (i.e. poor absorption, metabolic susceptibility, rapid excretion, limited access to the central nervous system, and high plasma protein binding). This is mainly due to high molecular weight, poor water solubility, and susceptible peptide linkages. In the following examples, we will find that potent PIs were discovered relatively quickly, but that these had a high peptide character. Subsequent work was then needed to reduce the peptide character of these compounds in order to retain high activity, whilst gaining acceptable levels of oral bioavailability and half-life.

Clinically useful PIs are generally less well absorbed from the gastrointestinal tract than reverse transcriptase inhibitors, and are also susceptible to first pass metabolic reactions involving the cytochrome P450 isozyme (CYP3A4). This metabolism can result in drug–drug interactions with many of the other drugs given to AIDS patients to combat opportunistic diseases (e.g. rifabutin, ketoconazole, rifampin, and astemizole).

20.7.4.1 The HIV protease enzyme

The HIV protease enzyme (Fig. 20.15) is an example of an enzyme family called the **aspartyl proteases**—enzymes which catalyse the cleavage of peptide bonds and which contain an aspartic acid in the active site that is crucial to the catalytic mechanism. The enzyme is relatively small and can be obtained by synthesis. Alternatively, it can be cloned and expressed in fast-growing cells then purified in large quantities. The enzyme is crystallized with or without an inhibitor bound to the active site, making it an ideal target for structure-based drug design where

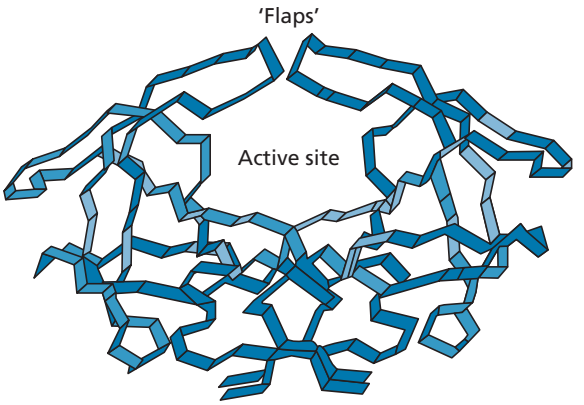


FIGURE 20.15 The HIV protease enzyme.

X-ray crystallographic studies of enzyme–inhibitor complexes allow the design of novel inhibitors.

The HIV protease enzyme is a symmetrical dimer made up of two identical protein units, each consisting of 99 amino acids. The active site is at the interface between the protein units and is also symmetrical with twofold rotational (C2) symmetry. The amino acids Asp-25, Thr-26, and Gly-27 from each monomer are located on the floor of the active site, and each monomer provides a flap to act as the ceiling. The enzyme has a broad substrate specificity and can cleave a variety of peptide bonds in viral polypeptides, but, crucially, it can cleave bonds between a proline residue and an aromatic residue (phenylalanine or tyrosine) (Fig. 20.16). The cleavage of a peptide bond next to proline is unusual and does not occur with mammalian proteases such as renin, pepsin, or cathepsin D, and so the chances of achieving selectivity against HIV protease over mammalian proteases are good. Moreover, the symmetrical nature of the viral enzyme and its active site is not present in mammalian proteases, again suggesting the possibility of drug selectivity.

There are eight binding subsites in the enzyme—four on each protein unit, located on either side of the catalytic region (Fig. 20.16). These subsites accept the amino acid side chains of the substrate and are numbered **S1–S4** on one side and **S1′–S4′** on the other side. The relevant side chains on the substrate are numbered **P1–P4** and **P1′–P4′** (Fig. 20.17). Peptide bonds in the substrate are also involved in hydrogen bonding interactions with the active site, as shown in Fig. 20.17. A water

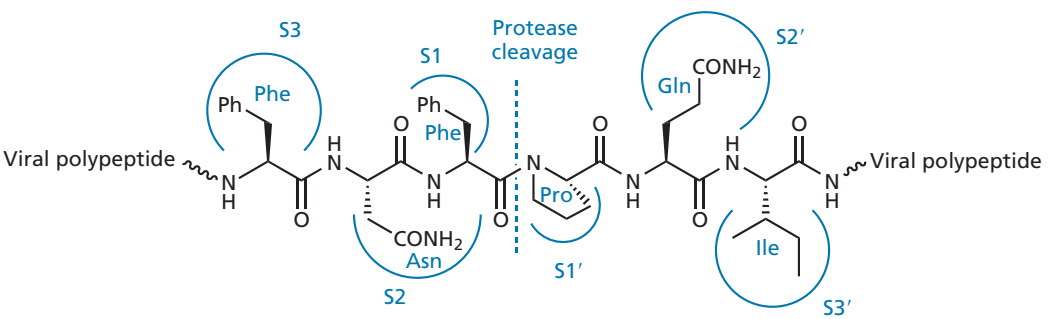


FIGURE 20.16 The aromatic–proline peptide bond that is cleaved by HIV protease (six of the eight binding subsites are shown).

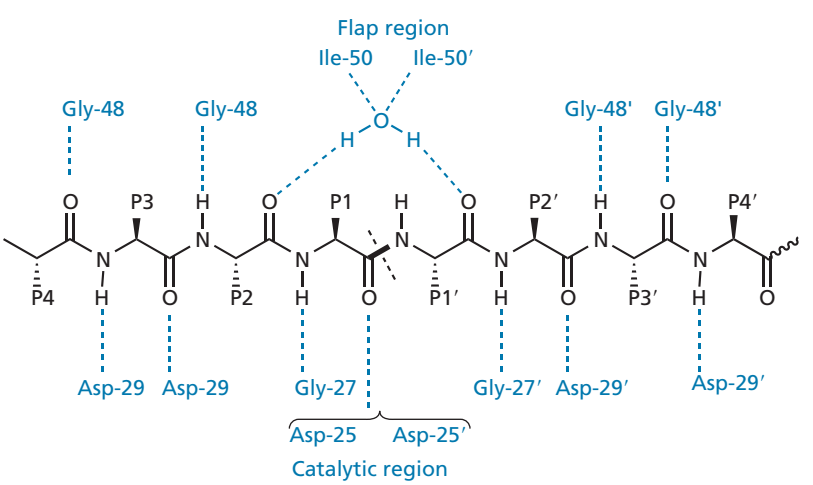


FIGURE 20.17 Interactions between the substrate's peptide backbone and the active site of HIV protease.

molecule is present in the active site which acts as a hydrogen bonding bridge to two isoleucine NH groups on the enzyme flaps. This hydrogen bonding network has the effect of closing the flaps over the active site once the substrate is bound.

There are two variants of HIV protease. The protease enzyme for HIV-2 shares 50% sequence identity with the protease enzyme for HIV-1. The greatest variation occurs outwith the active site and so inhibitors are found to bind similarly to both enzymes.

The aspartic acids Asp-25 and Asp-25′ on the floor of the active site are involved in the catalytic mechanism. Each of these residues is contributed by one of the protein subunits, and the carboxylate side chains interact with a bridging water molecule during the hydrolysis mechanism (Fig. 20.18).

20.7.4.2 Design of HIV protease inhibitors (PIs)

A similar hydrolytic mechanism to that shown in Fig. 20.18 takes place for a mammalian aspartyl protease called **renin**. This enzyme was studied extensively before the discovery of HIV protease, and a variety of renin inhibitors were designed as antihypertensive agents (section 7.4). These agents act as **transition-state inhibitors** and many of the strategies resulting from the development of renin inhibitors were adapted to the design of HIV PIs.

Transition-state inhibitors are designed to mimic the transition state of an enzyme-catalysed reaction. The advantage of this approach is that the transition state is likely to be bound to the active site more strongly than either the substrate or product. Therefore, inhibitors resembling the transition state are also likely to be bound more strongly. In the case of the HIV protease-catalysed reaction, the transition state resembles the tetrahedral intermediate shown in Fig. 20.18. As such structures are inherently unstable, it is necessary to design an inhibitor which contains a **transition-state isostere**. Such an isostere would have a tetrahedral centre to mimic the tetrahedral centre of the transition state, yet be stable to hydrolysis. Fortunately, several such isosteres had already been developed in the design of renin inhibitors (Fig. 20.19). Thus, a large number of structures were synthesized incorporating these isosteres, with the hydroxyethylamine isostere

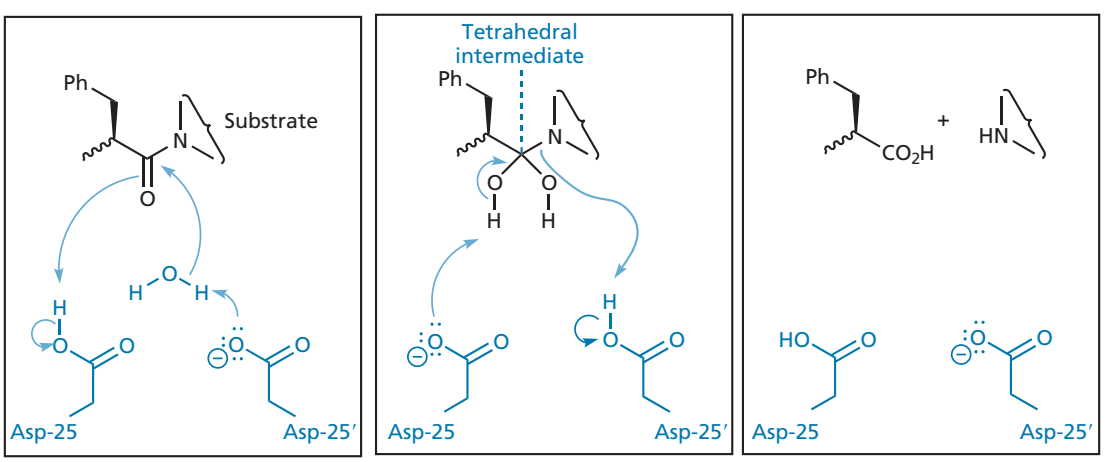


FIGURE 20.18 Mechanism of the reaction catalysed by HIV protease.

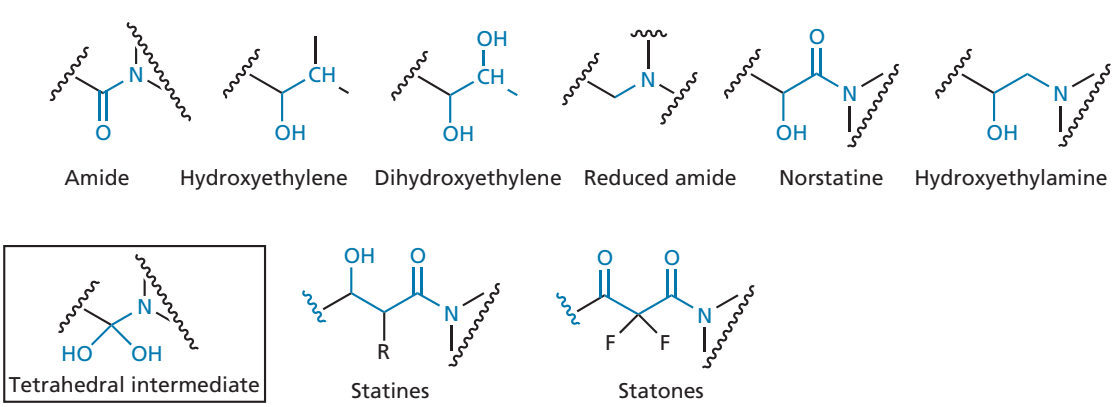


FIGURE 20.19 Transition-state isosteres.

proving particularly effective. This isostere has a hydroxyl group which mimics one of the hydroxyl groups of the tetrahedral intermediate and binds to the aspartate residues in the active site. The stereochemistry of this group is also important to activity, with the *R*-configuration generally being preferred. This preference is determined by the nature of the P1′ group that is present.

Having identified suitable transition-state isosteres, inhibitors were designed based on the enzyme's natural peptide substrates, as these contain amino acid residues which fit the eight subsites and allow a good binding interaction between the substrate and the enzyme. In theory, it might make sense to design inhibitors such that all eight subsites are filled to allow stronger interactions. However, this leads to structures with a high molecular weight and, consequently, poor oral bioavailability. Therefore, most of the PIs were designed to have a core unit spanning the S1 to S1′ subsites. Further substituents were then added at either end to fit into the S2/S3 and S2′/S3′ subsites. Early inhibitors, such as saquinavir (Fig. 20.21), have amino acid side chains that bind to most of the subsites from S3 to S3′. Unfortunately, these compounds have a large molecular weight and a high peptide character leading to poor pharmacokinetic properties. More recent inhibitors have been designed with increased aqueous solubility and oral bioavailability by using a variety of novel P2 and P2′ groups that reduce the molecular weight and peptide character of the compounds. The S2 and S2′ subsites of the protease enzyme appear to contain both polar (Asp-29, Asp-30) and hydrophobic (Val-32, Ile-50, Ile-84) amino acids, allowing the design of drugs that contain hydrophobic P2 groups which are also capable of hydrogen bonding. It has also been possible to design a P1 group that can span both the S1 and S3 subsites, allowing the removal of a P3 moiety, thus lowering the molecular weight. The P2 group is usually attached to P1 by an acyl link, because the carbonyl oxygen concerned acts as an important hydrogen bond acceptor to the bridging water molecule described previously (Fig. 20.17).

We shall now look at how these strategies were used to design individual PIs.

20.7.4.3 Saquinavir

Saquinavir was developed by Roche and, as the first PI to reach the market, it serves as the benchmark for all other PIs. The design of saquinavir started by considering a viral polypeptide substrate (pol, see section 20.7.2) and identifying a region of the polypeptide which contains a phenylalanine–proline peptide link. A pentapeptide sequence Leu–Asn–Phe–Pro–Ile was identified and served as the basis for inhibitor design. The peptide link normally hydrolysed in this sequence is between Phe and Pro, and so this link was replaced by a hydroxyethylamine transition-state isostere to give a structure which successfully inhibited the enzyme (Fig. 20.20). The amino acid side chains for Leu–Asn–Phe–Pro–Ile are retained in this structure and bind to the five subsites S3–S2′. Despite that, enzyme inhibition is relatively weak. The compound also has high molecular weight and peptide-like character, both of which are detrimental to oral bioavailability.

Consequently, the Roche team set out to identify a smaller inhibitor, starting from the simplest possible substrate for the enzyme—the dipeptide Phe–Pro (Fig. 20.21). The peptide link was replaced by the hydroxylamine transition-state isostere and the resulting *N*- and *C*-protected structure (I) was tested and found to have weak inhibitory activity. The inclusion of an asparagine group (structure II) to occupy the S2 subsite resulted in a 40-fold increase in activity, which meant that structure II was more active than the pentapeptide analogue (Fig. 20.20). This might seem an unexpected result as the latter occupies more binding subsites. However, it has been found that the crucial interaction of inhibitors is in the core region S2–S2′. If the addition of extra groups designed to bind to other subsites weakens the interaction to the core subsites, it can lead to an overall drop in activity. For example, the addition of leucine to structure II resulted in a drop in activity, despite the fact that leucine can occupy the S3 subsite.

Structure II was adopted as the new lead compound and the residues P1 and P2 were varied to find the optimum groups for the S1 and S2 subsites. As it turned out, the benzyl group and the asparagine side chain were already the optimum groups. An X-ray crystallographic study of the enzyme–inhibitor complex revealed that the protecting group (Z) occupied the S3 subsite, which proved to be a large hydrophobic pocket. Therefore, the protecting group was replaced with a larger quinoline ring system which could occupy the subsite more fully. This led to a sixfold increase in activity (structure III).

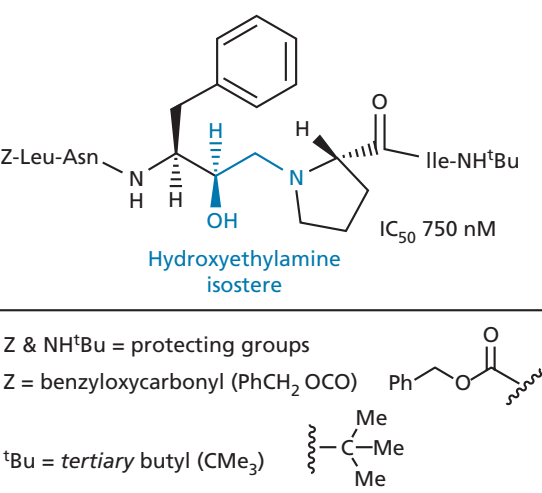


FIGURE 20.20 Pentapeptide analogue incorporating a hydroxyethylamine transition-state isostere.

Variations were also carried out on the carboxyl half of the molecule. Proline fits into the S1′ pocket, but it was possible to replace it with a bulkier decahydroisoquinoline ring system. The *t*-butyl ester protecting group was found to occupy the S2′ subsite and could be replaced by a *t*-butylamide group which proved more stable in animal studies. The resulting structure (saquinavir) showed a further 60-fold increase in activity. The *R*-stereochemistry of the transition-state hydroxyl group is essential. If the configuration is *S*, all activity is lost.

X-ray crystallography of the enzyme–saquinavir complex (Figs 20.21 and 20.22) demonstrated the following:

- the substituents on the drug occupy the five subsites S3–S2′;
- the *t*-butylamine nitrogen is positioned in such a way that further *N*-substituents would be incapable of reaching the S3′ subsite;
- there are hydrogen bonding interactions between the hydroxyl group of the hydroxyethylamine moiety and the catalytic aspartates (Asp-25 and Asp-25′);
- the carbonyl groups on either side of the transition-state isostere act as hydrogen bond acceptors to a bridging water molecule. The latter forms hydrogen bonds to the isoleucine groups in the enzyme's flap region in a similar manner to that shown in Fig. 20.17.

Saquinivir is still used clinically but suffers from poor oral bioavailability and susceptibility to drug resistance. Various efforts have been made to design simpler analogues of saquinavir which have lower molecular weight, less peptide character, and, consequently, better oral bioavailability.

Test your understanding and practise your molecular modelling with Exercise 20.2.

20.7.4.4 Ritonavir and lopinavir

Ritonavir was developed by Abbott Pharmaceuticals to take advantage of the symmetrical properties of the protease enzyme and its active site. Because the active site has

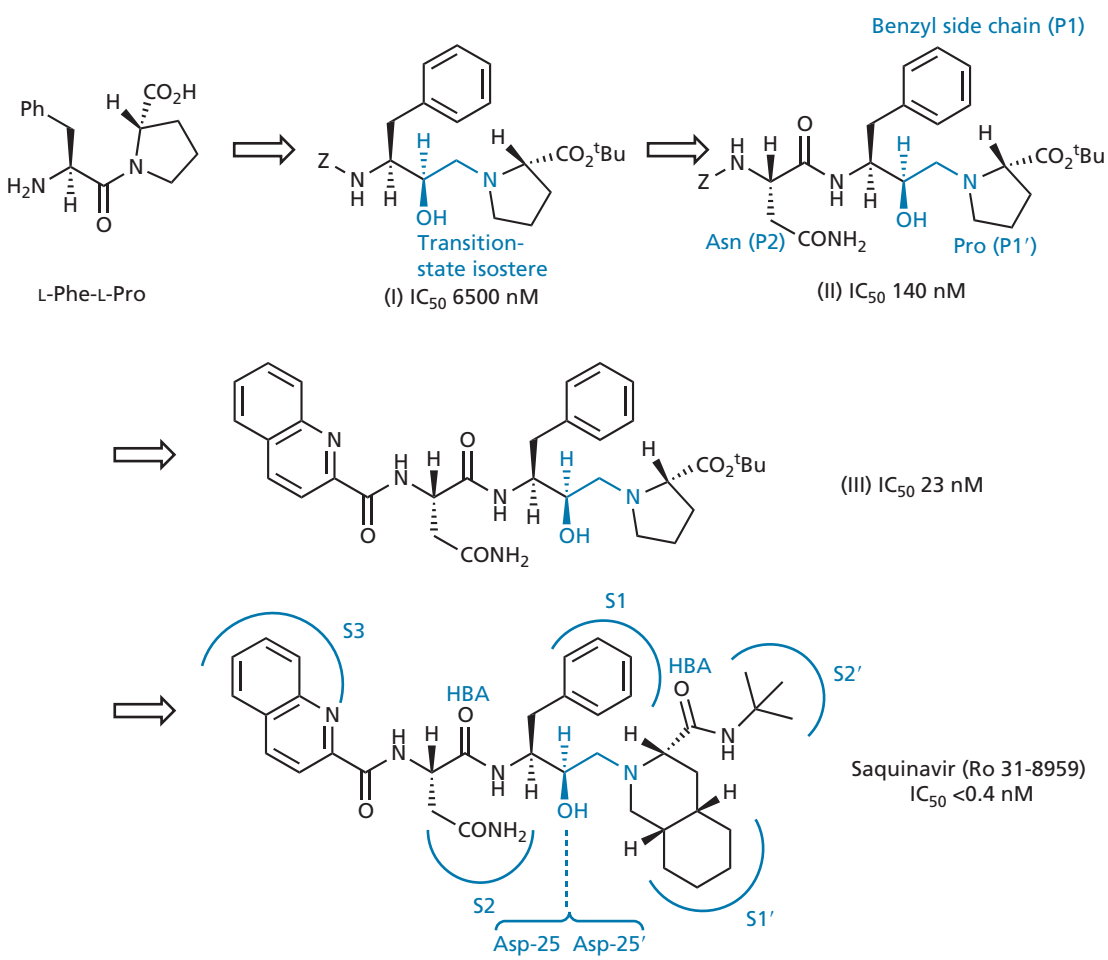


FIGURE 20.21 Development of saquinavir (Z = $PhCH_2OCO$).

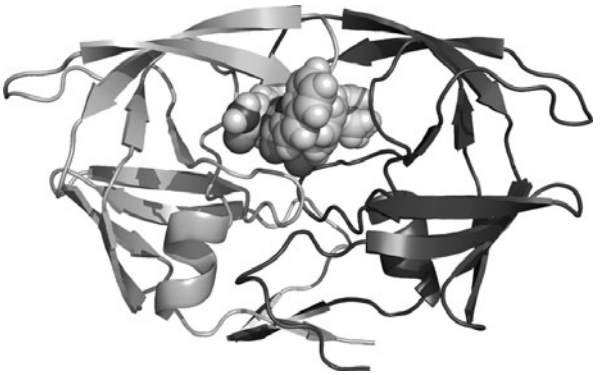


FIGURE 20.22 Saquinavir bound to the active site of HIV protease.

C2 symmetry, a substrate is capable of binding 'left to right' or 'right to left' as the binding subsites S1–S4 are identical to subsites S1′–S4′. This implies that it should be possible to design inhibitors that have C2 symmetry which could have several advantages. Firstly, symmetrical inhibitors should show greater selectivity for the viral protease over mammalian aspartyl proteases, as the active sites of the latter are not symmetrical. Secondly, symmetrical molecules might be less recognizable to peptidases, resulting in improved oral bioavailability. Thirdly, the development of saquinavir showed that a benzyl residue was the optimum binding group for the S1 subsite. As the S1′ subsite is identical to S1, a symmetrical inhibitor having benzyl groups fitting both S1 and S1′ subsites should bind more strongly and have improved activity. This argument could also be extended for the binding groups fitting the S2/S2′ subsites and so on (see Molecular modelling exercise 20.3).

As there was no lead compound having C2 symmetry to match the symmetry of the active site, it was necessary to design one. This was done by considering the tetrahedral reaction intermediate derived from the natural substrate. It was assumed that the axis of C2 symmetry for the active site passed through the reaction centre of this intermediate (Fig. 20.23). As the benzyl group was known to be optimum for binding to the S1 subsite, the left-hand portion of the molecule was retained and the right-hand portion was deleted. The left-hand moiety was then rotated such that two benzyl residues were present in the correct orientation for C2 symmetry. The resulting geminal diol is inherently unstable, so one of the alcohols was removed leading to the simplest target alcohol (I; R = H). In order to check whether this target molecule would match the C2 symmetry of the active site when bound, a molecular modelling experiment was carried out whereby the inhibitor was constructed in the active site. The results of this analysis were favourable and so the target alcohol was synthesized. Although it had no antiviral activity, it did show weak activity as an enzyme inhibitor, which meant that it could serve as a lead compound for further development. This represented a success for *de novo* techniques in the design of a lead compound (section 17.15).

The next stage was to extend the molecule to take advantage of the S2 and S2′ subsites. A variety of structures were synthesized and tested, revealing vastly improved enzyme inhibition when valine was added, and further improvement when the valines had *N*-protecting groups (A 74704; Fig. 20.24). A 74704 also showed *in vitro* activity against HIV and was resistant to proteolytic degradation. The structure was co-crystallized

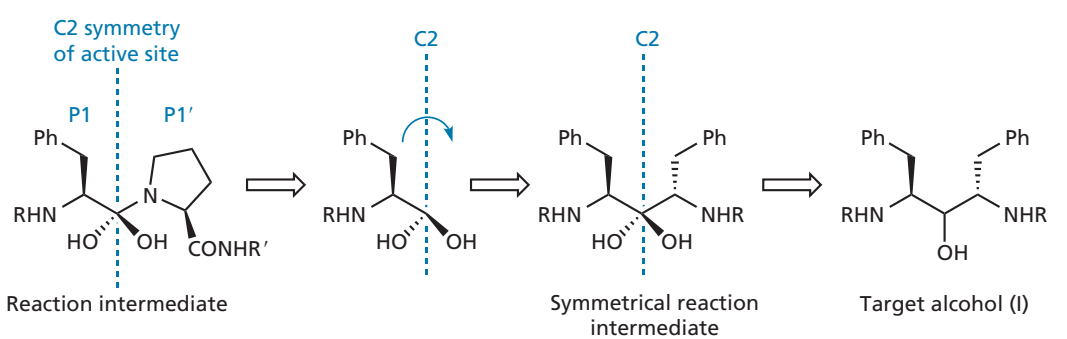


FIGURE 20.23 *De novo* design of a symmetrical lead compound acting as an inhibitor.

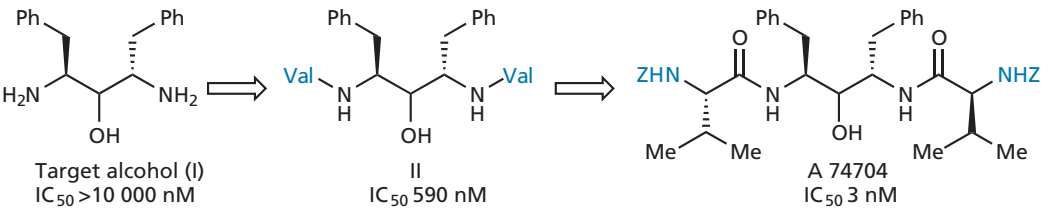


FIGURE 20.24 Development of A 74704 (Z = $PhCH_2OCO$).

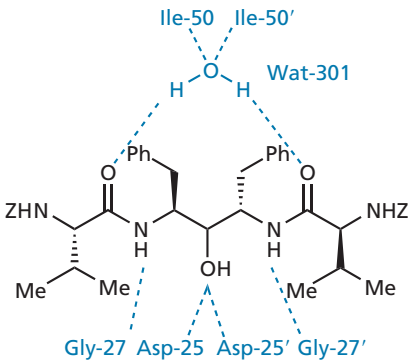


FIGURE 20.25 Binding interactions between the backbone of A 74704 and the active site of HIV protease (Z = $PhCH_2OCO$).

with recombinant protease enzyme and studied by X-ray crystallography to reveal a symmetrical pattern of hydrogen bonding between the inhibitor and the enzyme (Fig. 20.25). It was also found that a water molecule (Wat-301) still acted as a hydrogen bonding bridge between the carbonyl groups of P2 and P2′, and the NH groups of Ile-50 and Ile-50′ on the flaps of the enzyme. The C2 symmetry axes of the inhibitor and the active site passed within 0.2 Å of each other and deviated by an angle of only 6°, demonstrating the validity of the design philosophy.

Further analysis of the crystal structure suggested that the NH groups on the inhibitor were binding to Gly-27 and Gly-27′, but were too close to each other to allow optimum hydrogen bonding. To address this, it was decided to design symmetrical inhibitors where the relevant NH groups would be separated by an extra bond. In order to achieve this, the axis of C2 symmetry was placed through the centre of the susceptible bond. Accordingly, the design process was repeated to generate the diol shown in Figure 20.26 as a possible lead compound.

Diol structures analogous to the alcohols described previously were synthesized and tested. Curiously, it was found that the absolute configuration of the diol centres had little effect on activity and that the activity of the diols was generally better than the corresponding alcohols. For example, the diol equivalent of A 74704 (Fig. 20.27) had a 10-fold better level of activity. Unfortunately, this compound had poor water solubility, indicating that its polarity should be increased. A crystal structure of the enzyme–inhibitor complex revealed that the terminal portions of the molecule were exposed to solvation, which meant that more polar groups could be added at those positions without affecting binding. Consequently, the terminal phenyl groups were replaced by more polar pyridine rings. The urethane groups near the terminals were also replaced by urea groups, leading to A 77003 with improved water solubility. Unfortunately, the oral bioavailability was still unsatisfactory and so the structure entered clinical trials as an intravenous antiviral agent.

Modelling studies of how A 77003 might bind to the active site suggested two possible binding modes: one where each of the diol hydroxyl groups formed symmetrical hydrogen bonds to each of the aspartate residues, and one where only one of the hydroxyl groups hydrogen-bonded to both aspartate groups. To investigate this further, X-ray crystallography was carried out on the enzyme–inhibitor complex, revealing that asymmetric binding was taking place, whereby the (*R*)-OH took

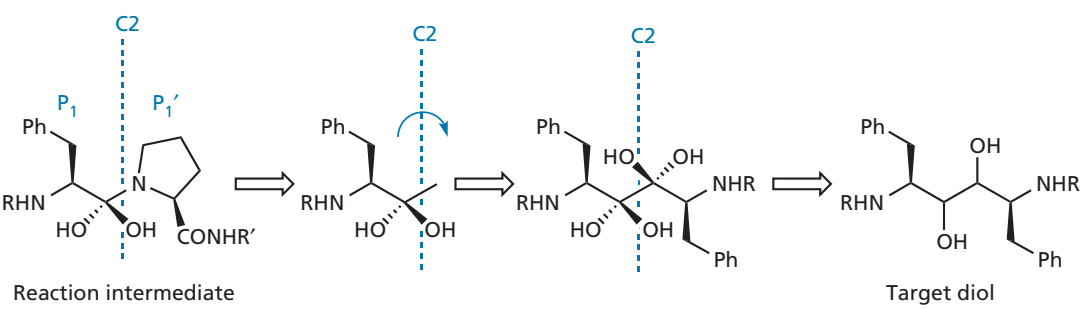


FIGURE 20.26 *De novo* design of a symmetrical diol inhibitor.

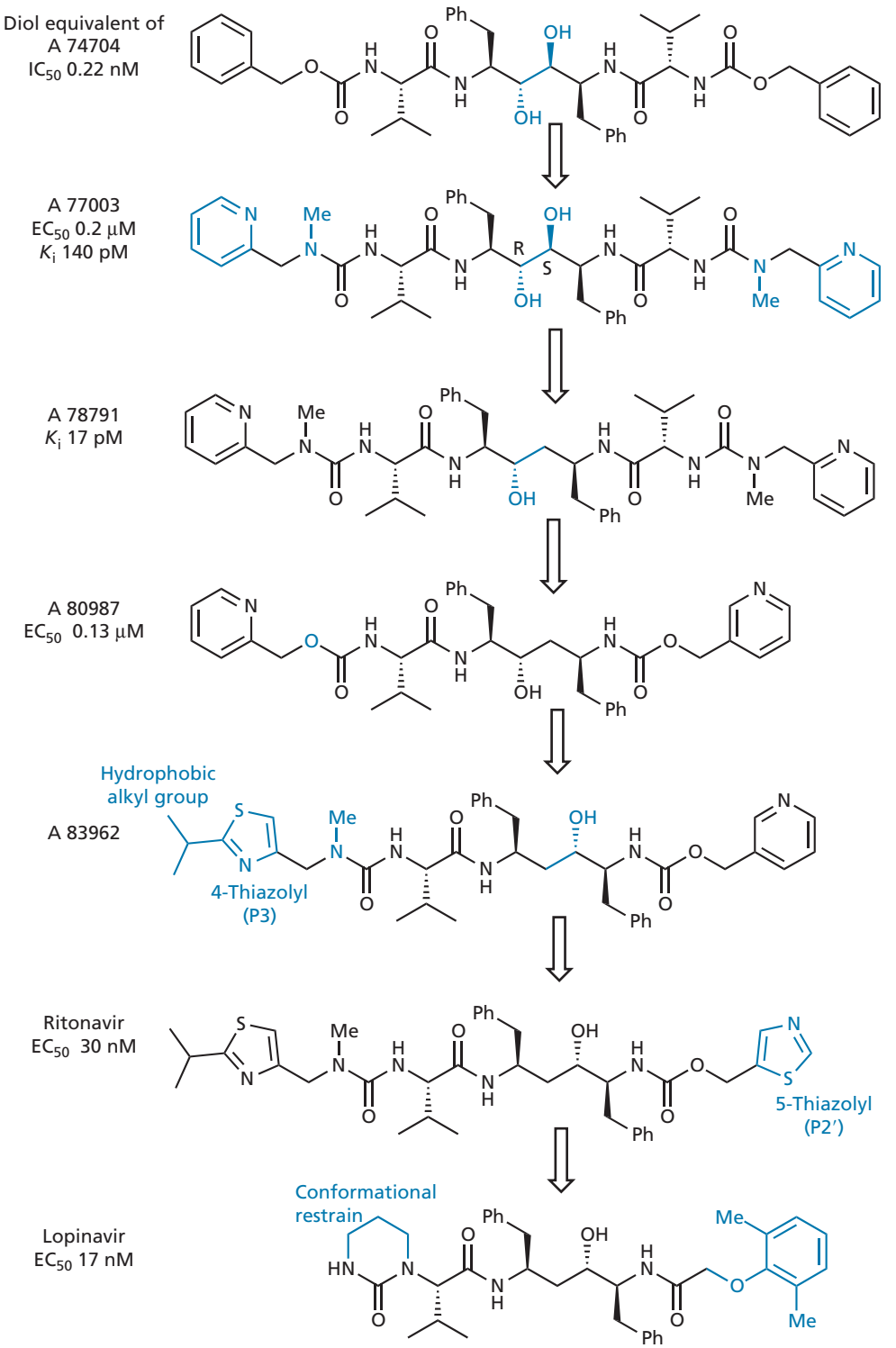


FIGURE 20.27 Development of ritonavir (ABT 538) and lopinavir (ABT 378).

part in hydrogen bonding with both aspartate residues, and the (*S*)-OH was only able to form a single hydrogen bonding interaction. This analysis also showed that the increased separation of the amide NHs failed to improve the geometry of the hydrogen bonding interactions with Gly-27 and Gly-27′. Thus, the improved activity of the diols over the alcohols was caused by reasons other than those proposed. Results such as these are not totally

unexpected when carrying out *de novo* design, as flexible molecules often bind differently from the manner predicted. The better activity for the diols may, in fact, be due to better binding of the P′ groups to the S′ subsites.

The fact that the (*S*)-hydroxyl group makes only one hydrogen bonding interaction suggested that it might be worth removing it, as the energy gained from only one hydrogen bonding interaction might be less than the energy required to desolvate the hydroxyl group before binding. This led to A 78791, which had improved activity and was shown by X-ray crystallography to bind in the same manner as A 77003.

A study was then carried out to investigate what effect variations of molecular size, aqueous solubility, and hydrogen bonding would have on the pharmacokinetics and activity of these agents. This led to A 80987, where the P2′ valine was removed, and the urea groups near the ends were replaced by urethane groups. In general, it was found that the presence of *N*-methylureas was good for water solubility and bioavailability, whereas the presence of urethanes (or carbamates) was good for plasma half-life and overall potency. Thus, it was possible to fine-tune these properties by a suitable choice of group at either end of the molecule.

Despite fewer binding interactions, A 80987 retained activity and had improved oral bioavailability. However, it had a relatively short plasma lifetime, was bound strongly to plasma proteins, and it was difficult to maintain therapeutically high levels. Metabolic studies then showed that A 80987 was *N*-oxidized at either or both pyridine rings and that the resulting metabolites were excreted mainly in the bile. In an attempt to counter this, various design strategies were carried out. Firstly, alkyl groups were placed on the pyridine ring at the vacant position *ortho* to the nitrogen. These were intended to act as a steric shield, but proved ineffective in preventing metabolism. It was then proposed that metabolism might be reduced if the pyridine rings were less electron-rich, and so methoxy or amino substituents were added as electron-withdrawing groups. However, this, too, failed to prevent metabolism. Finally, the pyridine ring at P3 was replaced by a variety of heterocycles in an attempt to find a different ring system which would act as a bioisostere, but which would be less susceptible to metabolism. The best results were obtained using the more electron-deficient 4-thiazolyl ring. Although water solubility decreased, it could be restored by reintroducing an *N*-methylurea group in place of one of the urethanes. Further improvements in activity were obtained by placing hydrophobic alkyl groups at the 2-position of the thiazole ring (P3) and by subsequently altering the position of the hydroxyl group in the transition-state isostere. This led to A 83962, which showed an eightfold increase in potency over A 80987.

Attention now turned to the pyridine group at P2′ which was replaced by a 5-thiazolyl group to give ritonavir, which had good activity and oral bioavailability. The good activity observed indicated that a hydrogen bonding interaction was taking place between the thiazolyl N and Asp-30 (specifically the NH of the peptide backbone). This matched a similar hydrogen bonding interaction involving the pyridine N in A 80987. The improved bioavailability is, principally, a result of better metabolic stability (20 times more stable than A 80987) and it was possible to get therapeutic plasma levels of the drug lasting 24 hours following oral administration.

Resistant strains of the virus have developed when ritonavir is used on its own. These arise from a mutation of valine at position 82 of the enzyme to alanine, threonine, or phenylalanine. X-ray crystallography shows that there is an important hydrophobic interaction between the isopropyl substituent on the P3 thiazolyl group of ritonavir and the isopropyl side chain of Val-82 which is lost as a result of this mutation. Further drug development led to lopinavir (Fig. 20.27) where the P3 thiazolyl group was removed and a cyclic urea group was incorporated to introduce conformational constraint. This allowed enhanced hydrogen bonding interactions with the S2 subsite, which balanced out the loss of binding due to the removal of the thiazolyl group. As this structure does not have any interactions with Val-82, it is active against the ritonavir-resistant strain.

20.7.4.5 Indinavir

The design of indinavir included an interesting hybridization strategy (Fig. 20.28). Merck had designed a potent PI that included a hydroxyethylene transition-state isostere (L 685,434). Unfortunately, it suffered from poor bioavailability and liver toxicity. At this point, the Merck workers concluded that it might be possible to take advantage of the symmetrical nature of the active site. As the S and S′ subsites are equivalent, it should be possible to combine half of one PI with half of another to give a structurally distinct hybrid inhibitor. A modelling study was carried out to check the hypothesis and the Merck team decided to combine the P′ half of L 685434 with the P′ half of saquinavir. The P′ moiety of saquinavir was chosen for its solubility enhancing potential and the P′ moiety of L 685434 is attractive for its lack of peptide character. The resulting hybrid structure (L 704,486) was less active as an inhibitor, but was still potent. Moreover, the presence of the decahydroisoquinoline ring system resulted in better water solubility and oral bioavailability (15%), as intended.

Further modifications were aimed at improving binding interactions, aqueous solubility, and oral bioavailability. The decahydroisoquinoline ring was replaced

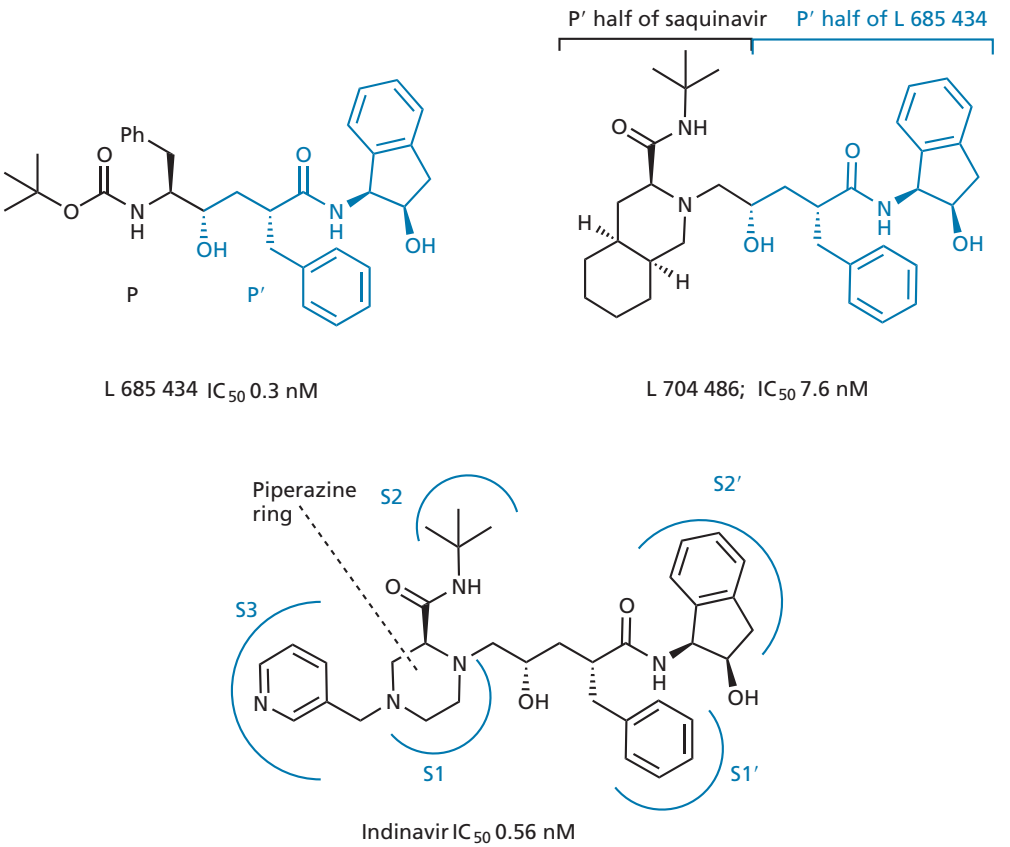


FIGURE 20.28 Development of indinavir.

by a piperazine ring, the additional nitrogen helping to improve aqueous solubility and oral bioavailability. A pyridine substituent was then added to access the S3 subsite and to improve binding. This resulted in indinavir, which reached the market in 1996. (See Molecular modelling exercise 20.4.)

For additional material see Web article 7: the design of indinavir.

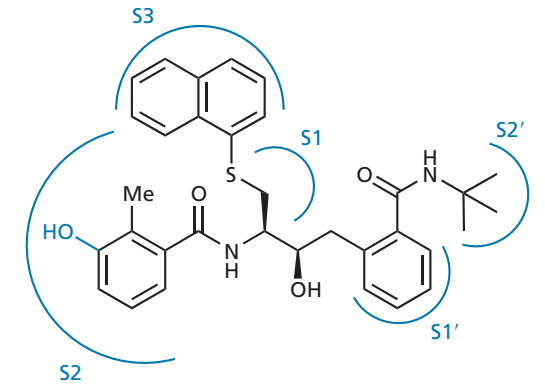
20.7.4.6 Nelfinavir

The development of nelfinavir was based on work carried out by the Lilly company, aimed at reducing the molecular weight and peptide character of PIs. Structure-based drug design had been used to develop AG1254 (Fig. 20.29), which contains an extended substituent at P1 capable of spanning and binding to both the S1 and S3 subsites of the enzyme. This did away with the need for a separate P3 group and allowed the design of compounds with a lower molecular weight. They also designed a new P2 group to replace an asparagine residue which had been present in their lead compound. This group was designed to bind effectively to the S2 subsite and, as it was different from any natural amino acid residue, the peptide character of the compound was reduced. Unfortunately, the antiviral activity of AG 1254 was not sufficiently high and the compound had poor aqueous solubility (see Molecular modelling exercise 20.3).

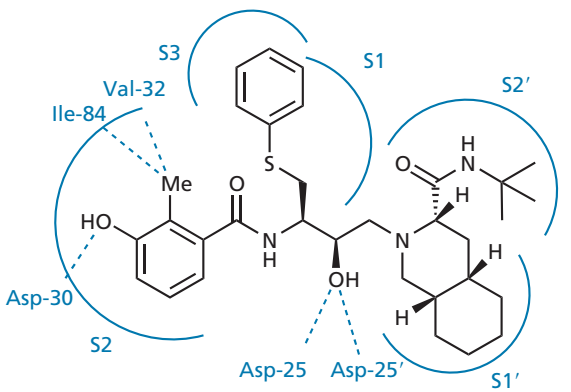
The company decided to switch direction and see what effect their newly designed substituents would have if they were incorporated into saquinavir and this led ultimately to nelfinavir. A crystal structure of nelfinavir bound to the enzyme showed that the molecule is bound in an extended conformation where the binding interactions involving the molecular backbone are similar to saquinavir. A tightly bound water molecule serves as a hydrogen bonding bridge between the two amide carbonyls of the inhibitor and the flap region of the enzyme, in a similar manner to other enzyme–inhibitor complexes. The crystal structure also showed that the *S*-phenyl group resides mainly in the S1 site and extends partially into the S3 site. The substituted benzamide group occupies the S2 pocket with the methyl substituent interacting with valine and isoleucine through van der Waals interactions, while the phenol interacts with Asp-30 through hydrogen bonding (see Web article 18: Nalfinavir (Viracepyt)).

20.7.4.7 Palinavir

Palinavir (Fig. 20.30) is a highly potent and specific inhibitor of HIV-1 and HIV-2 proteases. The left-hand or P half of the molecule is similar to saquinavir and the



AG 1254
K_i 3 nm



Nelfinavir
K_i 2.0 nM EC_{50} 0.008–0.02 μM

FIGURE 20.29 AG1254 and nelfinavir.

molecule contains the same hydroxyethylamine transition-state mimic. The right-hand (P′) side is different and was designed using the same kind of extension strategy used in nelfinavir. In this case, the P1′ substituent was extended to occupy the S1′ and S3′ subsites. This was achieved by replacing the original proline group at P1′ with 4-hydroxypipecolinic acid and adding a pyridine-containing substituent to access the S3′ subsite.

The crystal structure of the enzyme–inhibitor complex shows that the binding pockets S3–S3′ are all occupied. Two carbonyl groups interact via the bridging water molecule to the isoleucines in the enzyme flaps. The hydroxyl group interacts with both catalytic aspartate residues. Finally, the oxygen and NH atoms of all the amides are capable of hydrogen bonding to complementary groups in the active site. Work is currently in progress to simplify palinavir by introducing a single group that will span two binding subsites, thus allowing the removal of the P3 binding group.

20.7.4.8 Amprenavir and darunavir

Amprenavir (Fig. 20.31) was designed by Vertex Pharmaceuticals as a non-peptide-like PI using saquinavir as the lead compound. Saquinavir suffers from having a high molecular weight and a high peptide character, both of which are detrimental to oral bioavailability. Therefore, it was decided to design a simpler analogue with a lower molecular weight and less peptide character, but which retained good activity. Firstly, the decahydroisoquinoline group in saquinavir was replaced by an isobutyl sulfonamide group to give structure I. This has the advantage of reducing the number of asymmetric centres from six to three, allowing easier synthesis of analogues. Further simplification and reduction of peptide character was carried out by replacing the P2 and P3 groups with a tetrahydrofuran (THF) carbamate, which had previously been found by Merck to be a good binding group for the S2 subsite. Finally, an amino group was introduced on the phenylsulphonamide group to increase water solubility and to enhance oral absorption. Fosamprenavir is a phosphate prodrug for amprenavir.

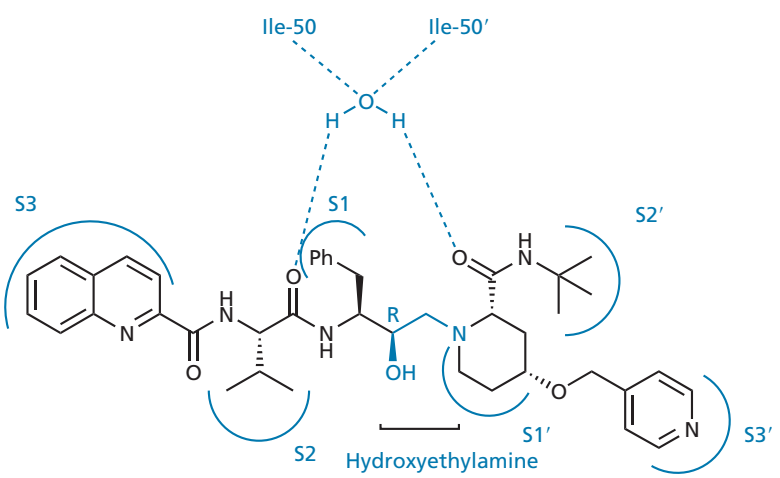


FIGURE 20.30 Palinavir and binding interactions.

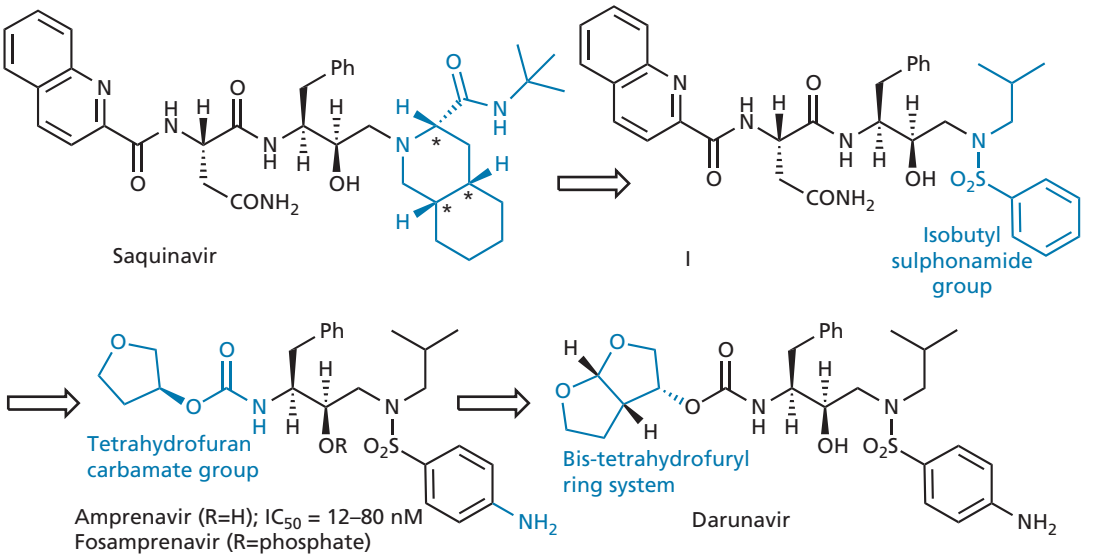


FIGURE 20.31 Development of amprenavir and darunavir.

Further work has shown that a fused bis-tetrahydrofuryl ring system is an even better binding group for the hydrophobic S2 pocket than a single THF ring because it fills the pocket more completely and forms hydrogen bonding interactions between the ring oxygens and the peptide backbone of the enzyme. As these interactions are with the protein backbone rather than amino acid side chains, mutations are less likely to lead to drug resistance. Darunavir is a second-generation PI which contains this feature, but there are several other compounds currently being studied.

20.7.4.9 Atazanavir

Atazanavir (Fig. 20.32) was approved in June 2003 as the first once-daily HIV-1 PI to be used as part of a combination therapy. It is similar to the early compounds leading towards ritonavir. Current research is looking at the possibility of using a deuterium-labelled analogue of atazanavir which is expected to have a slower rate of metabolism and excretion, and an increased half-life (see also section 14.2.4).

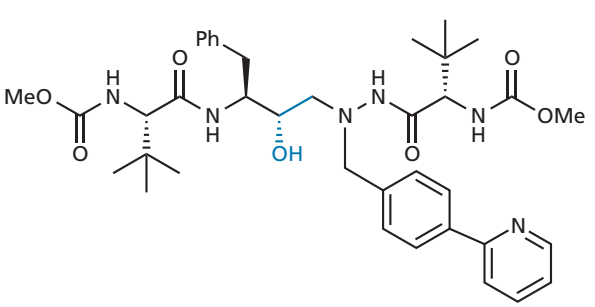


FIGURE 20.32 Atazanavir.

20.7.4.10 Tipranavir

Tipranavir (Fig. 20.33) is an example of a PI that was designed from a non-peptide lead compound. High throughput screening of 5000 structurally diverse compounds led to the discovery that the anticoagulant **warfarin** was a weak PI with antiviral activity. Various warfarin analogues were then tested leading to the discovery that **phenprocoumon** (Fig. 20.33) was a more potent competitive enzyme inhibitor with weak antiviral activity. Both these structures are used therapeutically for other purposes and have high oral bioavailability. Therefore, they served as promising lead compounds for non-peptide-like antiviral agents with good oral bioavailability.

A crystal structure of the enzyme–inhibitor complex was determined showing that the 4-OH group could form hydrogen bonds with the catalytic aspartate residues, while the two lactone oxygens could form hydrogen bonds directly to the isoleucine groups (Ile-50 and Ile-50′) in the enzyme flaps. Unlike all the previous PIs, there was no bridging water involved in this interaction. Therefore, these compounds represented a new class of inhibitors with a novel pharmacophore of hydrogen bonding interactions. The crystal structure also showed that the ethyl and phenyl groups fitted the S1 and S2 subsites, respectively, while the benzene ring of the coumarin ring system fitted the S1 subsite. Phenprocoumon was used as the lead compound for further development (see Web article 19) and resulted in the discovery of tipranavir.

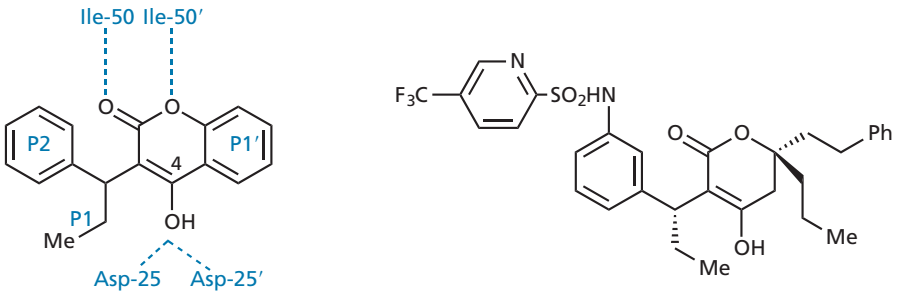


FIGURE 20.33 Phenprocoumon and tipranavir.

20.7.4.11 Alternative design strategies for antiviral drugs targeting the HIV protease enzyme

An alternative approach to inhibiting the protease enzyme would be to prevent its formation in the first place. Studies are in progress to design protein–protein binding inhibitors that will prevent the association of the two protein subunits that make it up (section 10.5).

Another interesting approach is to design prodrugs of toxic compounds that are only activated by HIV protease. The prodrugs would contain a moiety that acts as a substrate for HIV-protease, such that the toxin is only released in HIV-infected cells. The toxin would then attack cellular targets and eliminate those cells.

20.7.5 Inhibitors of other targets

Antisense agents are being developed to block the production of the HIV protein **Tat**, which is needed for the transcription of other HIV genes. **Trecovirsen** (Fig. 20.34) is a phosphorothioate oligonucleotide containing 25 nucleotides and has been designed to hybridize with the mRNA derived from the HIV gene *gag* to

BOX 20.4 Clinical aspects of protease inhibitors (PIs)

PIs are an important component of the drug cocktail used to treat HIV. Care has to be taken when administering the agents to haemophiliacs and diabetics as the agents can increase the risk of bleeding, and lower blood sugar levels.

Saquinivir was the first PI to reach the market in 1995. It shows a 100-fold selectivity for both HIV-1 and HIV-2 proteases over human proteases. Approximately 45% of patients develop clinical resistance to the drug over a 1-year period, but resistance can be delayed if it is given in combination with reverse transcriptase inhibitors. The oral bioavailability of saquinavir is only 4% in animal studies, although this is improved if the drug is taken with meals. The compound is also highly bound to plasma proteins (98%). As a result, the drug has to be taken in high doses to maintain therapeutically high plasma levels. A curious problem related to saquinavir is that plasma levels are lowered if the patient takes garlic.

Ritonavir reached the market in 1996. It is active against both HIV-1 and HIV-2 proteases and shows selectivity for HIV proteases over mammalian proteases. Despite the fact that ritonavir is highly plasma bound (99%), and has a high molecular weight and peptide-like nature, it has better bioavailability than many other PIs. This is a result of greater stability to drug metabolism, and it is possible to get therapeutic plasma levels of the drug which last 24 hours following oral administration. The metabolic stability of the agent is a result of the drug's ability to act as a potent inhibitor of the cytochrome P450 enzyme CYP3A4, which means that it shuts down its own metabolism. Care has to be taken when drugs affected by CYP3A4 are taken alongside ritonavir, and doses of the latter should be adjusted accordingly. However, ritonavir's ability to inhibit CYP3A4 is useful when it is used alongside other PIs which are normally metabolized by this enzyme (e.g. saquinavir, indinavir, nelfinavir, and amprenavir). As ritonavir inhibits CYP3A4, the lifetime and plasma levels of other PIs can be increased. For this reason, it is often administered in small doses alongside other PIs. If it is intended to be used as an anti-HIV drug in its own right, it is administered with nucleoside reverse transcriptase inhibitors (NRTIs).

(*Continued*)

BOX 20.4 Clinical aspects of protease inhibitors (PIs) (*Continued*)

Lopinavir is active against ritonavir-resistant strains of HIV, and is administered with ritonavir as a single capsule combination called Kaletra. Each capsule contains 133 mg of lopinavir and 33 mg of ritonavir with the latter serving as a cytochrome P450 inhibitor to increase the levels of lopinavir present in the blood supply.

Indinavir has better oral bioavailability than saquinavir and is less highly bound to plasma proteins (60%). It is usually administered alongside NRTIs such as didanosine.

Nelfinavir was marketed in 1997 and is used as part of a four-drug combination therapy. Like indinavir and ritonavir, nelfinavir is more potent than saquinavir because of its better pharmacokinetic profile. Compared with saquinavir, it has a lower molecular weight and log*P*, and an enhanced aqueous solubility, resulting in enhanced oral bioavailability. It can inhibit the metabolic enzyme CYP3A4 and thus affects the plasma levels of other drugs metabolized by this enzyme. It is 98% bound to plasma proteins.

Amprenavir was licensed to GlaxoWellcome and was approved in 1999. It is reasonably specific for the viral protease relative to mammalian proteases and is about 90% protein bound. It has good oral bioavailability (40–70% in animal studies). **Fosamprenavir** is a phosphate prodrug for amprenavir, and was approved by the US FDA and the European Medicines Agency (EMA) in 2003 and 2004 respectively. The prodrug acts as a slow-release version of amprenavir, reducing the number of pills required. It is usually administered with ritonavir. **Darunavir** is a second-generation PI developed by Tibotec, and was approved by the US FDA in 2006 as the first treatment of drug-resistant HIV. It is usually administered with ritonavir.

Atazanavir was approved in June 2003 as the first once-daily HIV-1 PI to be used as part of a combination therapy. It is usually administered with ritonavir.

Tipranavir is used to treat HIV infections that are resistant to other PIs. However, there have been cases of life-threatening liver toxicity.

prevent its translation into HIV proteins. It was withdrawn from clinical trials owing to toxicity but a similar oligonucleotide (**GEM92**) with increased stability is currently undergoing clinical trials.

Other agents under study for the treatment of HIV include **integrase inhibitors**, and **cell entry inhibitors**. Blocking entry of a virus into a host cell is particularly desirable, because it is so early in the life cycle. **Enfuvirtide** was approved in March 2003 as the first member of a new class of **fusion inhibitors**. It is a polypeptide consisting of 36 amino acids which matches the *C*-terminal end of the viral protein **gp41**. It works by forming an α-helix and binding to a group of three similar α-helices belonging to the gp41 protein. This association prevents the process by which the virus enters the host cell. In order to bring about fusion, the gp41 protein anchors the virus to the cell membrane of the host cell. It then undergoes a conformational change where it builds a grouping of six helices using the three already present as the focus for that grouping (Fig. 20.35). This pulls the membranes of the virion and the host cell together to permit fusion. By binding to the group of three helices, enfuvirtide blocks formation of the required hexamer and prevents fusion.

d(*P*-thio)(T-C-T-T-C-C-T-C-T-C-T-C-T-A-C-C-C-A-C-G-C-T-C-T-C)

FIGURE 20.34 Trecovirsen.

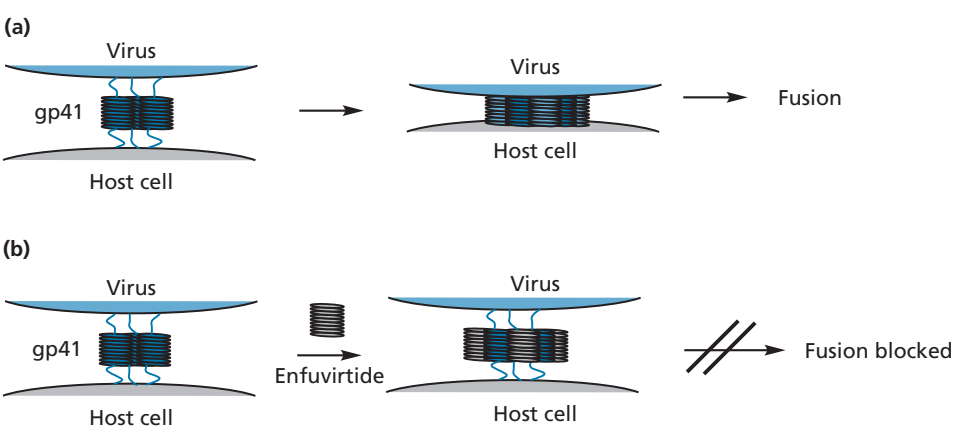


FIGURE 20.35 (**a**) Normal mechanism of fusion. (**b**) Enfuvirtide acting as a fusion inhibitor.

The manufacture of enfuvirtide involves 106 steps, which makes it expensive and may limit its use. A smaller compound (**BMS 378806**) is being investigated which binds to **gp120** and prevents the initial binding of the virus to **CD4** on the cell surface.

N-**Butyldeoxynojirimycin** (Fig. 20.36) is a carbohydrate that inhibits **glycosidases**—enzymes that catalyse the trimming of carbohydrate moieties linked to viral proteins. If this process is inhibited, too many carbohydrate groups end up attached to a protein, resulting in the protein adopting a different conformation. It is thought that the gp120 protein is affected in this way and cannot be peeled away as described in section 20.7.1 to reveal the gp41 protein.

Bicyclams such as **JM 3100** (Fig. 20.36) block the **CCR5 chemokine receptor** and are under investigation as drugs which will prevent membrane fusion and cell entry.

Maraviroc (Fig. 20.37) was approved as a CCR5 antagonist in 2007 and is the first anti-HIV agent to act on a molecular target on the host cell rather than the virus. It was developed from a compound that had potent activity, but which blocked HERG ion channels (Box 12.3). Agents which block these channels often have toxic cardiac side effects and so a large number of analogues were synthesized to find a potent compound which did not block the HERG ion channels. Maraviroc was the result. It is an example of an agent that works by blocking protein–protein interactions between a viral protein and a host cell protein (section 10.5).

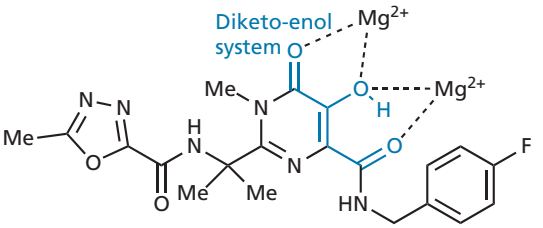
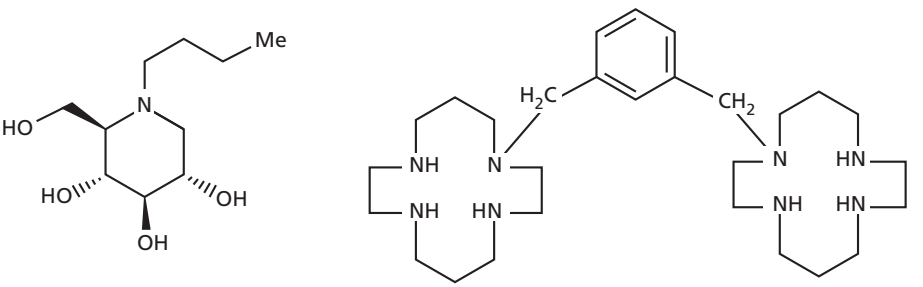


FIGURE 20.38 The integrase inhibitor raltegravir acting as a metal ion chelator.

The first integrase inhibitor to reach the market in 2007 was **raltegravir** (Fig. 20.38). The keto-enol system is important for activity as it acts as a chelating group for two magnesium ion cofactors in the enzyme's active site.

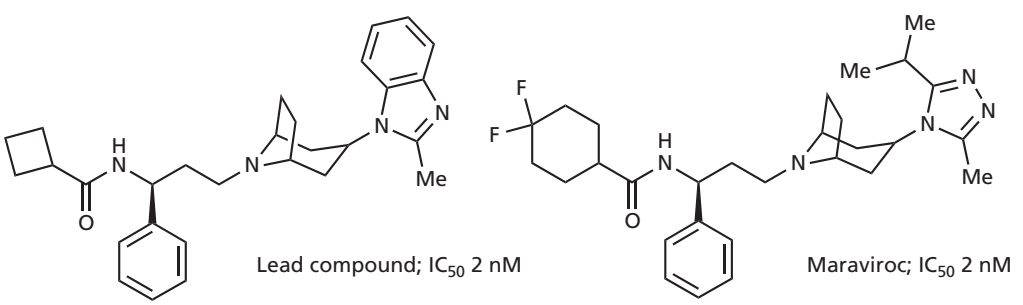
KEY POINTS

- HIV is a retrovirus containing RNA as its genetic material and is responsible for AIDS.
- The two main viral targets for anti-HIV drugs are the enzymes reverse transcriptase and protease. Combination therapy is the favoured treatment, but there is a need to develop drugs which are effective against a third target.



N-Butyldeoxynojirimycin JM 3100

FIGURE 20.36 Agents that inhibit cell entry.



Lead compound; IC_{50} 2 nM Maraviroc; IC_{50} 2 nM

FIGURE 20.37 Comparison of meraviroc and the lead compound from which it was developed.

- The potency and safety demands for anti-HIV drugs are high, as they are likely to be used for the lifetime of the patient.
- Reverse transcriptase is a DNA polymerase which catalyses the conversion of ssRNA to dsDNA. No such biochemical process occurs in normal cells.
- Nucleoside reverse transcriptase inhibitors (NRTIs) are prodrugs that are converted by cellular enzymes to active triphosphates which act as enzyme inhibitors and chain terminators.
- Non-nucleoside reverse transcriptase inhibitors (NNRTIs) act as enzyme inhibitors by binding to an allosteric binding site.
- The protease enzyme is a symmetrical dimeric structure consisting of two identical protein subunits. An aspartic acid residue from each subunit is involved in the catalytic mechanism.
- The protease enzyme is distinct from mammalian proteases in being symmetrical and being able to catalyse the cleavage of peptide bonds between proline and aromatic amino acids.
- Protease inhibitors (PIs) are designed to act as transition-state inhibitors. They contain a transition-state isostere which is tetrahedral but stable to hydrolysis. Suitable substituents are added to fill various binding pockets usually occupied by the amino acid side chains of polypeptide substrates.
- To obtain an orally active PI, it is important to maximize the binding interactions with the enzyme, while minimizing the molecular weight and peptide character of the molecule.
- Cell fusion inhibitors have been developed, one of which has reached the market.
- The first integrase inhibitor has been approved for the market.

20.8 Antiviral drugs acting against RNA viruses: flu virus

20.8.1 Structure and life cycle of the influenza virus

Influenza (or flu) is an airborne, respiratory disease caused by an RNA virus which infects the epithelial cells of the upper respiratory tract. It is a major cause of mortality, especially among the elderly or among patients with weak immune systems. The most serious pandemic occurred in 1918 with the death of at least 20 million people worldwide caused by the Spanish flu virus. Epidemics then occurred in 1957 (Asian flu), 1968 (Hong Kong flu), and 1977 (Russian flu). Despite the names given to these flus, it is likely that they all derived from China where families live in close proximity to poultry and pigs, increasing the chances of viral infections crossing from one species to another.* In 1997, there was an outbreak of flu in Hong Kong which killed 6 out of the 18 people infected. This was contained by slaughtering infected chickens, ducks, and geese which had been the source of the problem. If action had not been swift, it is possible that this flu variant could have become a pandemic and wiped out 30% of the world's population. This emphasizes the need for effective antiviral therapies to combat flu.

The nucleocapsid of the flu virus contains (–) ssRNA and a viral enzyme called **RNA polymerase** (see Fig. 20.1). Surrounding the nucleocapsid there is a membranous envelope derived from host cells which contains two viral glycoproteins called **neuraminidase** (NA) and **haemagglutinin** (HA). The latter acquired its name because it can bind virions to red blood cells and cause haemagglutination. The NA and HA glycoproteins are spike-like objects which project about 10 nm from the surface and are crucial to the infectious process.

In order to reach the epithelial host cells of the upper respiratory tract, the virus has to negotiate a layer of protective mucus and it is thought that the viral protein NA is instrumental in achieving this. The mucosal secretions are rich in glycoproteins and glycolipids which bear a terminal sugar substituent called **sialic acid** (also called ***N*-acetylneuraminic acid**). NA (also called **sialidase**) is an enzyme which cleaves sialic acid from these glycoproteins and glycolipids (Fig. 20.39), thus degrading the mucus layer and allowing the virus to reach the surface of epithelial cells.

Once the virus reaches the epithelial cell, adsorption takes place whereby the virus binds to cellular glycoconjugates that are present in the host cell membrane, and which have a terminal sialic acid moiety. The viral protein HA is crucial to this process. Like NA, it recognizes sialic acid but, instead of catalysing the cleavage of sialic acid from the glycoconjugate, HA binds to it (Fig. 20.40). Once the virion has been adsorbed, the cell membrane bulges inwards taking the virion with it to form a vesicle called an **endosome**—a process called **receptor mediated endocytosis**. The pH in the endosome then decreases, causing HA in the virus envelope to undergo a dramatic conformational change whereby the hydrophobic ends of the protein spring outward and extend towards the endosomal membrane. After contact, fusion occurs and the RNA nucleocapsid is released into the cytoplasm of

* On the other hand, there has been a recent theory that the 1918 pandemic originated in army transit camps in France. The living conditions in these camps were similar to communities in China in the sense that large numbers of soldiers were camping in close proximity to pigs and poultry used as food stocks. The return of the forces to all parts of the globe following World War I could explain the rapid spread of the virus.

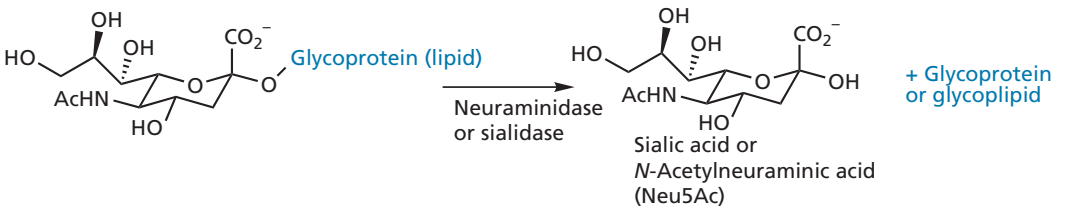


FIGURE 20.39 Action of neuraminidase (sialidase).

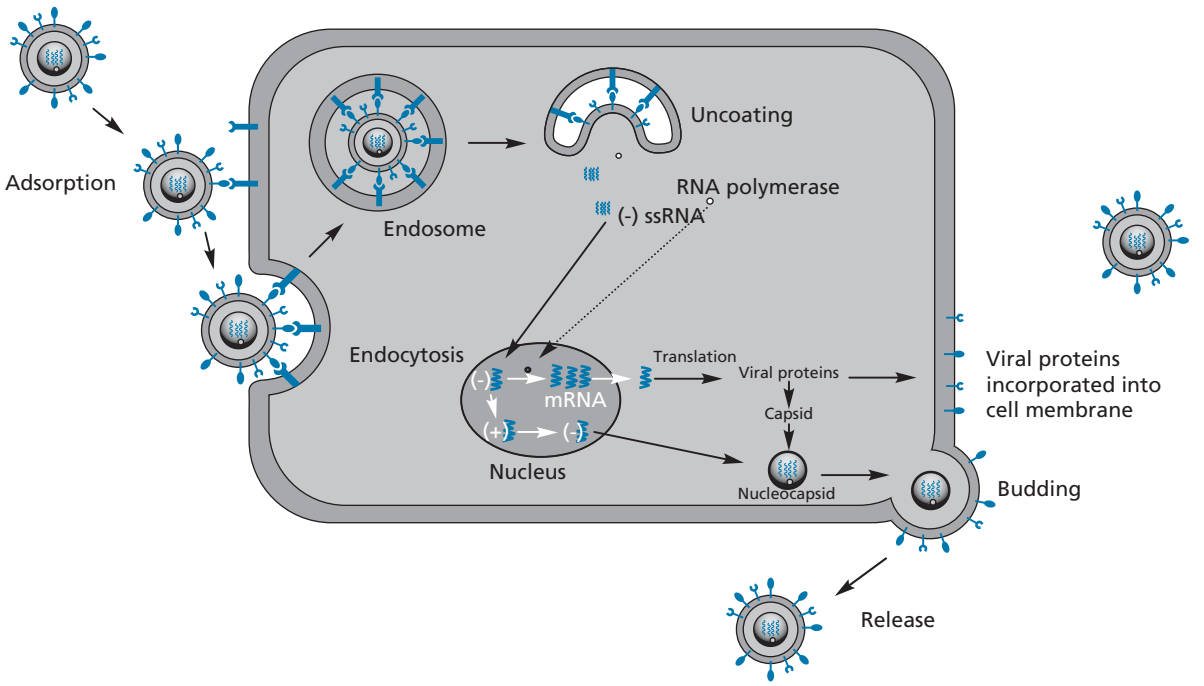


FIGURE 20.40 Life cycle of the influenza virus in a host epithelial cell.

the host cell. Disintegration of the nucleocapsid releases viral RNA and viral RNA polymerase, which both enter the cell nucleus.

Viral RNA polymerase now catalyses the copying of (−) viral RNA to produce (+) viral RNA, which departs the nucleus and acts as the mRNA required for the translation of viral proteins. Copies of (−) viral RNA are also produced in the nucleus, then exported into the cytoplasm.

Capsid proteins spontaneously self-assemble in the cytoplasm with incorporation of (−) RNA and newly produced RNA polymerase to form new nucleocapsids. Meanwhile, the freshly synthesized viral proteins HA and NA are incorporated into the membrane of the host cell. Newly formed nucleocapsids then move to the cell membrane and attach to the inner surface. HA and NA move through the cell membrane to concentrate at these areas and host cell proteins are excluded. Budding then takes place and a new virion is released. NA aids this release by hydrolysing any interactions that take place between HA on the virus and sialic acid conjugates on the host cell membrane.

There is an important balance between the rate of desialylation by NA (to aid the virion's departure from the host cell) and the rate of attachment by HA to sialylated glycoconjugates (to allow access to the cell). If NA was too active, it would hinder infection of the cell by destroying the receptors recognized by HA. However, if NA was too weak, the newly formed virions would remain adsorbed to the host cell after budding, preventing them from infecting other cells. It is noticeable that the amino acids present in the active site of NA are highly conserved, unlike amino acids elsewhere in the protein. This demonstrates the importance of the enzyme's level of activity.

As HA and NA are on the outer surface of the virion, they can act as antigens (i.e. molecules which can be potentially recognized by antibodies and the body's defence systems). In theory, it should be possible to prepare vaccines which will allow the body to gain immunity from the flu virus. Such vaccinations are available,

but they are not totally protective and they lose what protective effect they have with time. This is because the flu virus is adept at varying the amino acids present in HA and NA, thus making these antigens unrecognizable to the antibodies which originally recognized them—a process called **antigenic variation**. The reason it takes place can be traced back to the RNA polymerase enzyme, which is a relatively error-prone enzyme and means that the viral RNA which codes for HA and NA is not consistent. Variations in the code lead to changes in the amino acids present in NA and HA, which results in different types of flu virus based on the antigenic properties of their NA and HA. For example, there are nine antigenic variants of NA.

There are three groups of flu virus, classified as A, B, and C. Antigenic variation does not appear to take place with influenza C and occurs slowly with influenza B. With influenza A, however, variation occurs almost yearly. If the variation is small, it is called **antigenic drift**. If it is large, it is called **antigenic shift** and it is this that can lead to the more serious epidemics and pandemics. There are two influenza A virus subtypes which are epidemic in humans—those with H1N1 antigens and those with H3N2 antigens (where H and N stand for HA and NA respectively). A major aim in designing effective antiviral drugs is to find a drug that will be effective against the influenza A virus and remain effective despite antigenic variations. In general, vaccination is the preferred method of preventing flu, but antiviral drugs have a role in the prevention and treatment of flu when vaccination proves unsuccessful.

20.8.2 Ion channel disrupters: adamantanes

The adamantanes (Fig. 20.41) were discovered by random screening and are the earliest antiviral drugs used clinically against flu, decreasing the incidence of the disease by 50–70%. **Amantadine** and **rimantadine** (Fig. 20.41) are related adamantanes with similar mechanisms of action and can inhibit viral infection in two ways. At low concentration (<1 μg/ml), they inhibit the replication of influenza A viruses by blocking a viral ion channel protein called **matrix (M2) protein**. At higher concentrations (>50 μg/ml), the basic nature of the compounds becomes important and they buffer the pH of endosomes to prevent the acidic environment needed for HA to fuse the viral membrane with that of the endosome. These mechanisms inhibit penetration and uncoating of the virus.

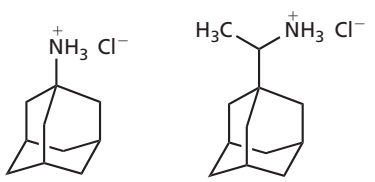


FIGURE 20.41 The adamantanes.

Unfortunately, the virus can mutate in the presence of amantadine to form resistant variants. Amantadine binds to a specific region of the M2 ion channel, and resistant variants have mutations which alter the width of the channel. Research carried out to find analogues which might still bind to these mutants proved unsuccessful. Work has also been carried out in an attempt to find an analogue which might affect the ion channel and pH levels at comparable concentrations. This has focused on secondary and tertiary amines with increased basicity, as well as alteration of the structure to reduce activity for the ion channel. The rationale is that resistant flu variants are less likely to be produced if the drug acts on two different targets at the same time. Rimantadine was approved in 1993 as a less toxic alternative to amantadine for the treatment of influenza A. Unfortunately, neither agent is effective against influenza B as this virus does not contain the matrix (M2) protein. Side effects are also a problem, possibly owing to effects on host cell ion channels.

20.8.3 Neuraminidase inhibitors

20.8.3.1 Structure and mechanism of neuraminidase

Since neuraminidase (NA) has two crucial roles in the infectious process (section 20.8.1), it is a promising target for potential antiviral agents. Indeed, a screening program for NA inhibitors was carried out as early as 1966, although without success. Following on from this, researchers set out to design a mechanism-based transition-state inhibitor. This work progressed slowly until the enzyme was isolated and its crystal structure studied by X-ray crystallography and molecular modelling.

NA is a mushroom-shaped tetrameric glycoprotein anchored to the viral membrane by a single hydrophobic sequence of some 29 amino acids. As a result, the enzyme can be split enzymatically from the surface and studied without loss of antigenic or enzymic activity. X-ray crystallographic studies have shown that the active site is a deep pocket located centrally on each protein subunit. There are two main types of the enzyme (corresponding to the influenza viruses A and B) and various subtypes. Due to the ease with which mutations occur, there is a wide diversity of amino acids making up the various types and subtypes of the enzyme. However, the 18 amino acids making up the active site itself are constant. As mentioned previously, the level of enzyme activity is

crucial to the infectious process and any variation that affects the active site is likely to affect the activity of the enzyme. This, in turn, will affect the infectious process adversely. As the active site remains constant, any inhibitor designed to fit it has a good chance of inhibiting all strains of the flu virus. Moreover, it has been observed that the active site is quite different in structure from the active sites of comparable bacterial or mammalian enzymes, so there is a strong possibility that inhibitors can be designed that are selective antiviral drugs.

The enzyme has been crystallized with sialic acid (the product of the enzyme-catalysed reaction) bound to the active site and the structure determined by X-ray crystallography. A molecular model of the complex was created which resembled the observed crystal structure as closely as possible. From this it was calculated that sialic acid was bound to the active site through a network of hydrogen bonds and ionic interactions as shown in Fig. 20.42.

The most important interactions involve the carboxylate ion of sialic acid, which is involved in ionic interactions and hydrogen bonds with three arginine residues, particularly with Arg-371. In order to achieve these interactions, the sialic acid has to be distorted from a stable chair conformation (where the carboxylate ion is in the axial position) to a less stable pseudo-boat conformation where the carboxylate ion is equatorial.

There are three other important binding regions or pockets within the active site. The glycerol side chain of sialic acid at C6 fills one of these pockets, interacting with glutamate residues and a water molecule by hydrogen bonding. The hydroxyl group at C-4 is situated in another binding pocket, interacting with a glutamate residue. Finally, the acetamido substituent at C5 fits into a hydrophobic pocket which is important for molecular recognition. This pocket includes the hydrophobic residues Trp-178 and Ile-222 which lie close to the methyl carbon (C-11) of sialic acid, as well as the hydrocarbon backbone of the glycerol side chain.

It was further established that the distorted pyranose ring binds to the floor of the active site cavity through its hydrophobic face. The glycosidic OH at C-2 is also shifted from its normal equatorial position to an axial position where it points out of the active site and can form a hydrogen bond to Asp-151, as well as an intramolecular hydrogen bond to the hydroxyl group at C-7.

Based on these results, a mechanism of hydrolysis was proposed which consists of four major steps (Fig. 20.43). The first step involves the binding of the substrate (sialoside), as described earlier. The second step involves proton donation from an activated water facilitated by the negatively charged Asp-151, and formation of an endocyclic sialosyl cation transition-state intermediate. Glu-277 is proposed to stabilize the developing positive charge on the glycosidic oxygen as the mechanism proceeds.

The final two steps of the mechanism are formation and release of sialic acid. Support for the proposed mechanism comes from kinetic isotope studies which indicate that it is an S_N1 nucleophilic substitution. NMR studies have also been carried out which indicate that sialic acid is released as the α-**anomer**. This is consistent with an S_N1 mechanism having a high degree of stereofacial selectivity. It is possible that expulsion of the product from the active site is favoured by mutarotation to the more stable β-anomer.

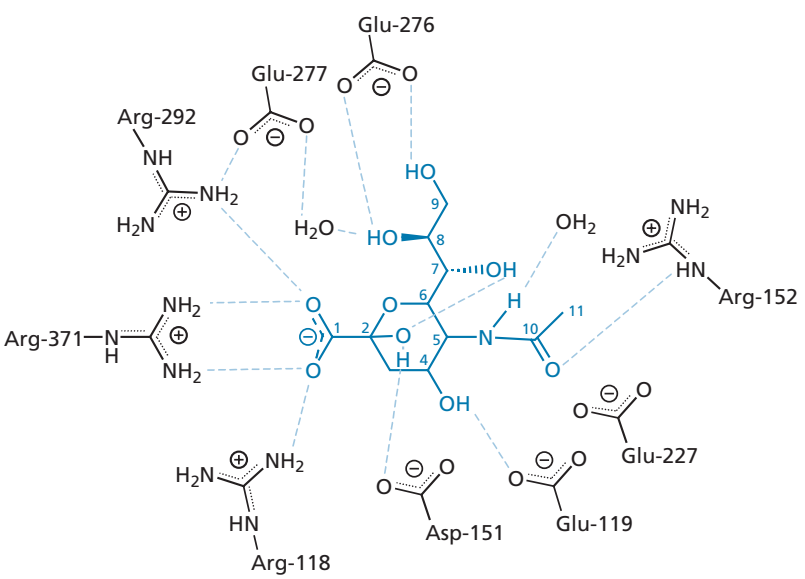


FIGURE 20.42 Hydrogen bonding interactions between sialic acid and the active site of neuraminidase.

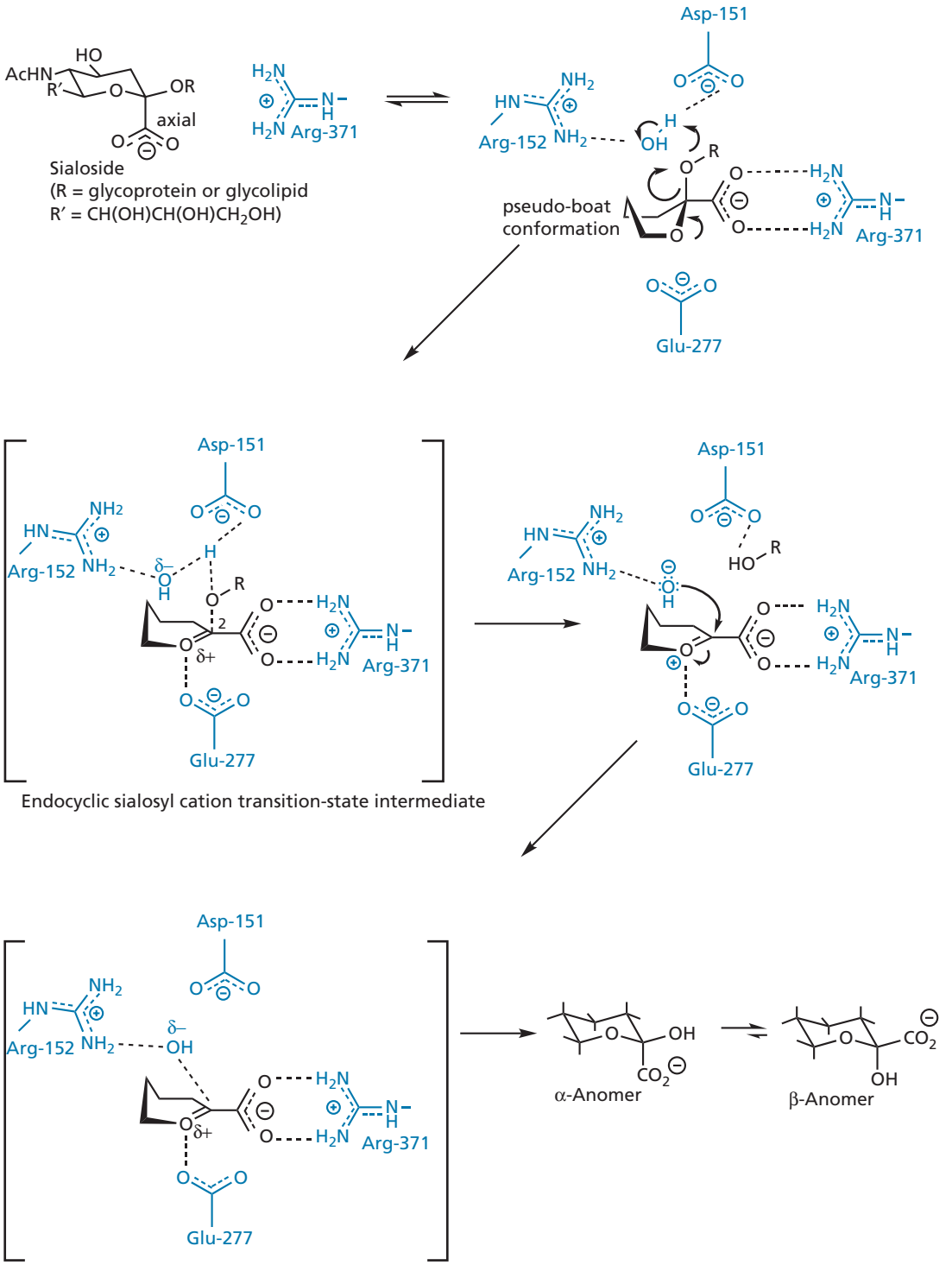


FIGURE 20.43 Proposed mechanism for the enzyme-catalysed hydrolysis of glycoconjugates to sialic acid (substituents are not shown during the mechanism for clarity).

Finally, site-directed mutagenesis studies have shown that the activity of the enzyme is lost if Arg-152 is replaced by lysine and Glu-277 is replaced by aspartate. These replacement amino acids contain similarly charged, but shorter, side chains. As a result, the charged side chains are unable to reach the required area of space in order to stabilize the intermediate.

20.8.3.2 Transition-state inhibitors: development of zanamivir (Relenza)

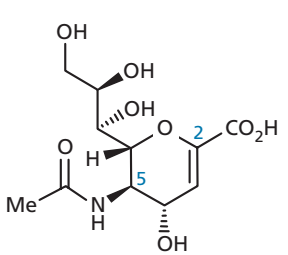
The transition state shown in Fig. 20.43 has a planar trigonal centre at C-2 and so sialic acid analogues containing a double bond between positions C-2 and C-3 were synthesized to achieve that same trigonal geometry at C-2. This resulted in the discovery of the inhibitor **2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid** (**Neu5Ac2en**) (Fig. 20.44). In order to achieve the required double bond, the hydroxyl group originally present at C-2 of sialic acid had to be omitted, which resulted in lower hydrogen bonding interactions with the active site. However, the inhibitor does not need to distort from a favourable chair shape in order to bind, and the energy saved by this more than compensates for the loss of one hydrogen bonding interaction. The inhibitor was crystallized with the enzyme and studied by X-ray crystallography and molecular modelling to show that the same binding interactions were taking place with the exception of the missing hydroxyl group at C-2. Unfortunately, this compound also inhibited bacterial and mammalian sialidases and could not be used therapeutically. Moreover, it was inactive *in vivo*.

The search for new inhibitors centred around the use of **GRID** molecular modelling software to evaluate likely binding regions within a model active site. This involved setting up a series of grid points within the active site and placing probe atoms at each point to measure interactions between the probe and amino acid residues (section 17.7.5). Different atomic probes were used to represent various functional groups. These included the oxygen of a carboxylate group, the nitrogen of an aminium ion, the oxygen of a hydroxyl group, and the carbon of a methyl group. Multi-atom probes were also used. These were positioned in the grid such that one atom of the probe was placed at each grid point in turn. Energy calculations were performed for all the atoms within the probe to give a total interaction energy at each grid point. Each time, the probe was rotated to find the orientation for the best hydrogen-bonding interaction.

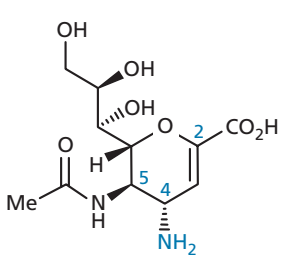
The most important result from these studies was the discovery that the binding region normally occupied by the 4-OH of sialic acid could also interact with an aminium or guanidinium ion. As a result, sialic acid analogues having an amino or guanidinyl group at C-4, instead of a hydroxyl group, were modelled in the active site to study the binding interactions and to check whether there was room for the groups to fit.

These modelling studies were favourable and so the relevant structures were synthesized and tested for activity. **4-Amino-Neu5Ac2en** (Fig. 20.44) contains the aminium group and was found to be more potent than Neu5Ac2en. Moreover, it was active in animal studies and showed selectivity against the viral enzyme, implying that the region of the active site which normally binds the 4-hydroxyl group of the substrate is different in the viral enzyme from comparable bacterial or mammalian enzymes. A crystal structure of the inhibitor bound to the enzyme confirmed the binding pattern predicted by the molecular modelling (Fig. 20.45).

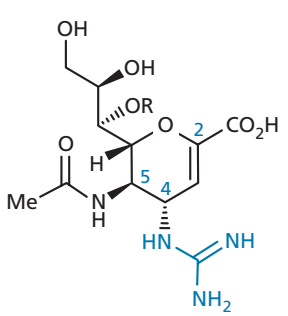
Molecular modelling studies had suggested that the larger guanidinium group would be capable of even greater hydrogen bonding interactions, as well as favourable van der Waals interactions. The relevant structure



Neu5Ac2en
K_i (M) 4×10^{-6}; IC_{50} 5–10 μM



4-Amino-Neu5Ac2en
K_i (M) 4×10^{-8}



Zanamivir (Relenza); R=H
K_i (M) 3×10^{-11}
Laninamivir (Inavir); R=Me

FIGURE 20.44 Transition-state inhibitors for the enzyme neuraminidase.

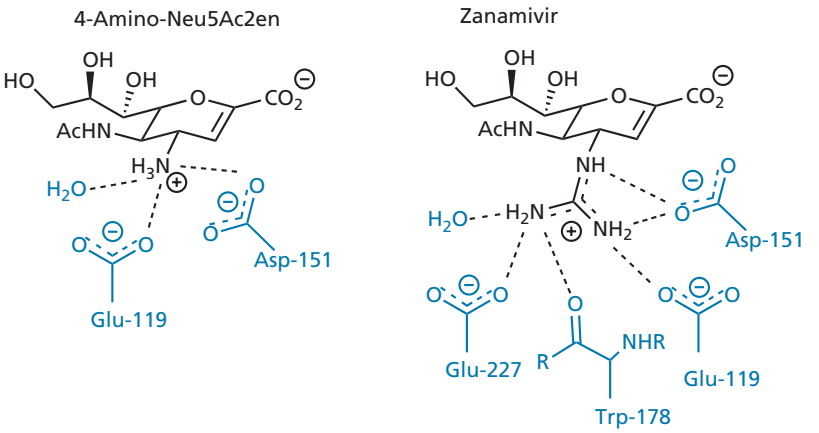


FIGURE 20.45 Binding interactions of aminium and guanidinium moieties at C-4 with the active site of neuraminidase.

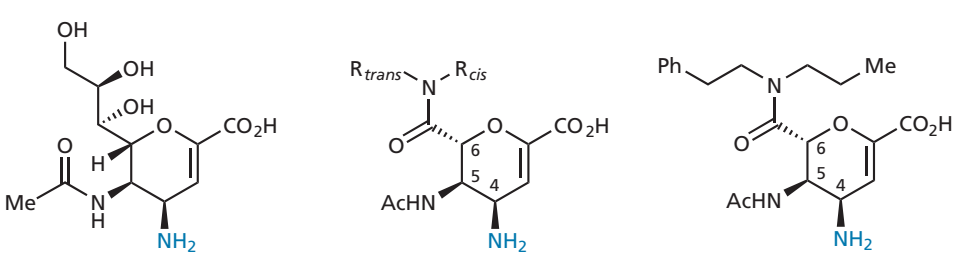
(**zanamivir**; Fig. 20.44) was, indeed, found to be a more potent inhibitor having a 100-fold increase in activity. X-ray crystallographic studies of the enzyme-inhibitor complex demonstrated the expected binding interactions (Fig. 20.45). Moreover, the larger guanidinium group was found to expel a water molecule from this binding pocket which is thought to contribute a beneficial entropic effect. Zanamivir is a slow-binding inhibitor with a high binding affinity to influenza A NA. It was approved by the US FDA in 1999 for the treatment of influenza A and B, and was marketed by Glaxo Wellcome and Biota. Unfortunately, the polar nature of the molecule means it has poor oral bioavailability (<5%), and it is administered by inhalation. **Laninamivir** (Fig. 20.44) is a closely related structure which was approved in Japan during 2010.

Following on from the success of these studies, **4-epi-amino-Neu5Ac2en** (Fig. 20.46) was synthesized to place the amino group in another binding region predicted by the GRID analysis. This structure proved to be a better inhibitor than Neu5Ac2en, but not as good as zanamivir. The pocket into which this amino group fits is small and there is no room for larger groups.

20.8.3.3 Transition-state inhibitors: 6-carboxamides

A problem with the inhibitors described in section 20.8.3.2 is their polar nature. The glycerol side chain is particularly polar and has important binding interactions with the active site. However, it was found that it could be replaced by a carboxamide side chain with retention of activity (Fig. 20.46).

A series of 6-carboxamide analogues was prepared to explore their structure–activity relationships. Secondary carboxamides (where R*cis* = H) showed similar weak inhibition against both A and B forms of the NA enzyme. Tertiary amides having an alkyl substituent at the *cis* position resulted in a pronounced improvement against the A form of the enzyme, with relatively little effect on the activity against the B form. Thus, tertiary amides showed a marked selectivity of 30–1000-fold for the A form of the enzyme. Good activity was related to a variety of different sized R_{trans} substituents larger than methyl, but the size of the R_{cis} group was more restricted and optimum activity was achieved when R_{cis} was ethyl or *n*-propyl.



4-Epi-amino-Neu5Ac2en
K_i (M) 3×10^{-7}

6-Carboxamides

I

FIGURE 20.46 4-Epi-amino-Neu5Ac2en and carboxamides.

The 4-guanidinium analogues are more active than corresponding 4-amino analogues but the improvement is slightly less than that observed for the glycerol series, especially where the 4-amino analogue is already highly active.

Crystal structures of the carboxamide (I in Fig. 20.46) bound to both enzymes A and B were determined by X-ray crystallography (Fig. 20.47). The dihydropyran portion of the carboxamide (I) binds to both the A and B forms of the enzyme in essentially the same manner as observed for zanamivir. The important binding interactions involve the carboxylate ion, the 4-amino group, and the 5-acetamido group—the latter occupying a hydrophobic pocket lined by Trp-178 and Ile-222.

However, there is a significant difference in the region occupied by the carboxamide side chain. In the sialic acid analogues, the glycerol side chain forms intermolecular hydrogen bonds to Glu-276. These interactions are not possible for the carboxamide side chain. Instead, the Glu-276 side chain changes conformation and forms a salt bridge with the guanidino side chain of Arg-224, and reveals a lipophilic pocket into which the R_{cis} *n*-propyl substituent can fit. The size of this pocket is optimal for an ethyl or propyl group which matches the structure–activity (SAR) results. The R_{trans} phenethyl group lies in an extended lipophilic cleft on the enzyme surface formed between Ile-222 and Ala-246. This region can accept a variety of substituents, again consistent with SAR results.

Comparison of the X-ray crystal structures of the native A and B enzymes shows close similarity of position and orientation of the conserved active site residues except in the region occupied normally by the glycerol side chain, particularly with regard to Glu-276. Zanamivir can bind to both A and B forms with little or no distortion of the native structures. Binding of the carboxamide (I) to the A form is associated with a change in torsion angles of the Glu-276 side chain such that the residue can form the salt bridge to Arg-224, but there is little distortion of the protein backbone in order to achieve this. In contrast, when the carboxamide binds to the B form of the enzyme, a significant distortion of the protein backbone is required before the salt bridge is formed. Distortion in the B enzyme structure also arises around the phenethyl substituent. This implies that binding of the carboxamide to the B form involves more energy expenditure than to A and this can explain the observed specificity.

Although none of the carboxamides studied reached the market, the information gained from crystal studies on the binding interactions proved relevant in the development of oseltamivir (next section).

20.8.3.4 Carbocyclic analogues: development of oseltamivir (Tamiflu)

The dihydropyran oxygen of Neu5Ac2en and related inhibitors have no important role to play in binding these structures to the active site of NA. Therefore, it should be possible to replace it with a methylene isostere to form carbocyclic analogues such as structure I in Fig. 20.48. This would have the advantage of removing a polar oxygen atom which would increase hydrophobicity and potentially increase oral bioavailability. Moreover, it would be possible to synthesize cyclohexene analogues, such as structure II, which more closely match the stereochemistry of the reaction's transition state than previous inhibitors (compare the reaction intermediate in Fig. 20.48 which can be viewed as a transition-state mimic). Such agents might be expected to bind more strongly and be more potent inhibitors.

Structures I and II were synthesized to test this theory, and it was discovered that structure II was 40 times more

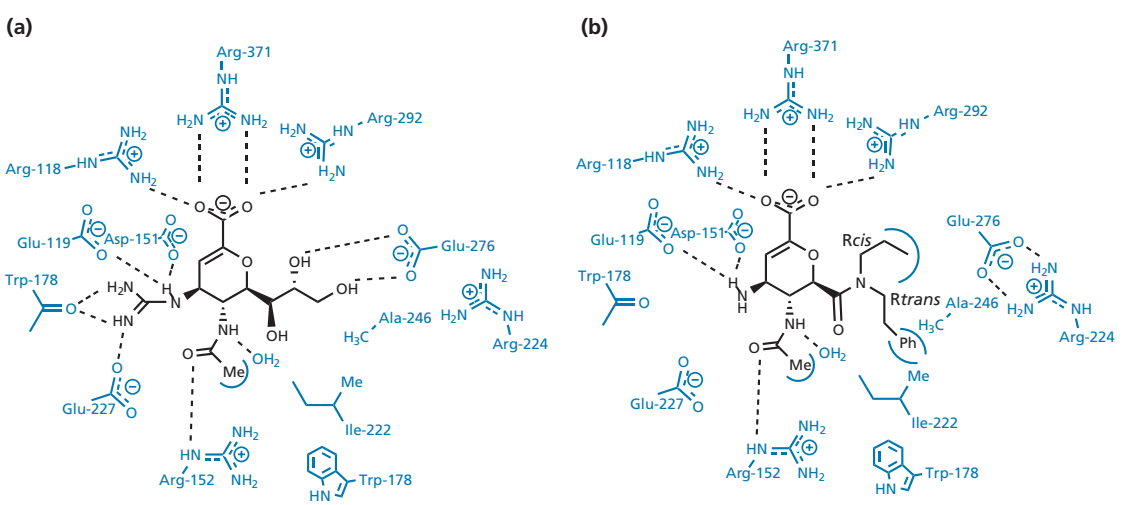


FIGURE 20.47 Binding interactions of zanamivir and carboxamides; (**a**) binding of zanamivir to the active site; (**b**) binding of carboxamide (I) to the active site.

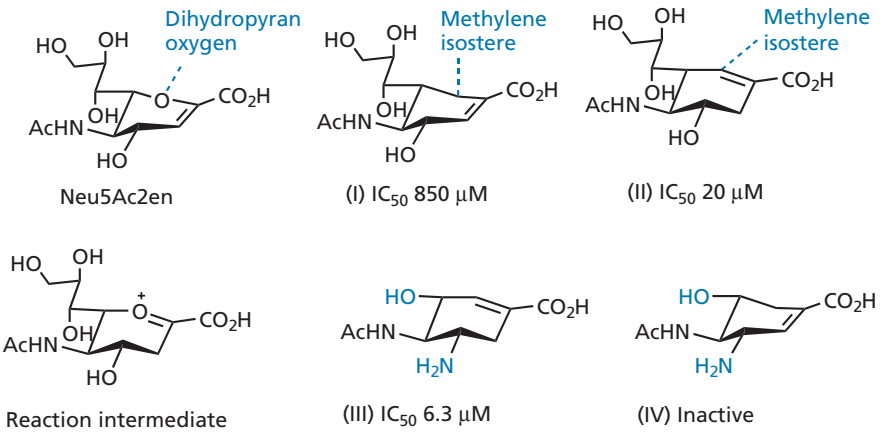


FIGURE 20.48 Comparison of Neu5Ac2en, reaction intermediate, and carbocyclic structures.

potent than structure I as an inhibitor. As the substituents are the same, this indicates that the conformation of the ring is crucial for inhibitory activity. Both structures have half-chair conformations, but these are different owing to the position of the double bond.

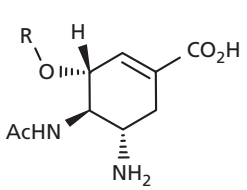
It was now planned to replace the hydroxyl group on the ring with an amino group to improve binding interactions (compare section 20.8.3.2), and to remove the glycerol side chain to reduce polarity. In its place, a hydroxyl group was introduced for two reasons. Firstly, the oxonium double bond in the transition state is highly polarized and electron deficient, whereas the double bond in the carbocyclic structures is electron rich. Introducing the hydroxyl substituent in place of the glycerol side chain means that the oxygen will have an inductive electron-withdrawing effect on the carbocyclic double bond and reduce its electron density. Secondly, adding the hydroxyl group meant that it would be possible to synthesize ether analogues which would allow the addition of hydrophobic groups to fill the binding pocket previously occupied by the glycerol side chain (compare section 20.8.3.3). The resultant structure III was synthesized and proved to be a potent inhibitor. In contrast, the isomer IV failed to show any inhibitory activity.

A series of alkoxy analogues of structure III was now synthesized in order to maximize hydrophobic interactions in the region of the active site previously occupied by the glycerol side chain (Fig. 20.49). For linear alkyl chains, potency increased as the carbon chain length increased from methyl to *n*-propyl. Beyond that, activity was relatively constant (150–300 nM) up to, and including, *n*-nonyl, after which activity dropped. Although longer chains than propyl increase hydrophobic interactions, there is a downside in that there is partial exposure of the side chain to water outside of the active site.

Branching of the optimal propyl group was investigated. There was no increase in activity when methyl branching was at the β-position, but the addition of a methyl group at the α-position increased activity by 20-fold. Introduction of an α-methyl group introduces an asymmetric centre, but both isomers were found to have similar activity indicating two separate hydrophobic pockets. The optimal side chain proved to be a pentyloxy side chain (R = $CH(Et)_2$).

The *N*-acetyl group is required for activity and there is a large drop in activity without it. The binding region for the *N*-acetyl group has limitations on the functionality and size of groups which it can accept. Any variations tend to reduce activity. This was also observed with sialic acid analogues.

Replacing the amino group with a guanidine group improves activity, as with the sialic acid series. However,



Linear chains (R)		Branched chains (R)		Miscellaneous chains (R)	
R	IC_{50} (μM)	R	IC_{50} (μM)	R	IC_{50} (μM)
Me	3.70	CH_2CHMe_2	0.200	CH_2OMe	2.00
Et	2.00	$CH(Me)CH_2CH_3$	0.010	$CH_2CH_2CF_3$	0.20
n-Pr	0.18	$CH(Et)_2$	0.001	$CH_2CH{=}CH_2$	2.20
n-Bu	0.30			cyclopentyl	0.02
				cyclohexyl	0.06
				phenyl	0.53

FIGURE 20.49 Alkoxy analogues.

the improvement in activity depends on the type of alkyl group present on the side chain, indicating that individual substituent contributions may not be purely additive.

The most potent of the above analogues was the pentyloxy derivative (GS 4071) (Fig. 20.50). This was co-crystallized with the enzyme and the complex was studied by X-ray crystallography, revealing that the alkoxy side chain makes several hydrophobic contacts in the region of the active site normally occupied by the glycerol side chain. In order to achieve this, the carboxylate group of Glu-276 is forced to orientate outwards from the hydrophobic pocket, as observed with the carboxamides. The overall gain in binding energy from these interactions appears to be substantial, as a guanidinium group is not required to achieve low nanomolar inhibition. Interactions elsewhere are similar to those observed with previous inhibitors.

Oseltamivir (Tamiflu) (Fig. 20.50) is the ethyl ester prodrug of GS 4071 and was approved in 1999 for the treatment of influenza A and B. The drug is marketed by Hoffman La Roche and Gilead Sciences. It is taken orally and is converted to GS4071 by esterases in the gastrointestinal tract. (See Molecular modelling exercise 20.5.)

20.8.3.5 Other ring systems

Work has been carried out to develop new NA inhibitors where different ring systems act as scaffolds for the important binding groups (Fig. 20.50).

The five-membered tetrahydrofuran (I) is known to inhibit NA with a potency similar to Neu5Ac2en. It has the same substituents as Neu5Ac2en, although their arrangement on the ring is very different. Nevertheless, a crystal structure of (I) bound to the enzyme shows that the important binding groups (carboxylate, glycerol, acetamido and C-4-OH) can fit into the required pockets. The central ring or scaffold is significantly displaced from the position occupied by the pyranose ring of Neu5Ac2en in order to allow this. This indicates that the position of the central ring is not crucial to activity and that the relative position of the four important binding groups is more important.

Five-membered carbocyclic rings have also been studied as possible scaffolds. Structure II (Fig. 20.50) was designed such that the guanidine group would fit the negatively-charged binding pocket described previously. A crystal structure of the inhibitor with the enzyme showed that the guanidine group occupies the desired pocket and displaces the water molecule originally present. It is involved in charge-based interactions with Asp-151, Glu-119, and Glu-227, analogous to zanamivir.

Modelling studies suggested that the addition of a butyl chain to the structure would allow van der Waals interactions with a small hydrophobic surface in the binding site. The target structure now has four asymmetric centres, and a synthetic route was used which controlled the configuration of two of these. As a result, four racemates, or eight isomers, were prepared as a mixture (Fig. 20.51). NA crystals were used to select the most active isomer of the mixture by soaking a crystal of the enzyme in the solution of isomers for a day and then collecting X-ray diffraction data from the crystal. This showed the active isomer to be structure I in Fig. 20.52. The structure binds to the active sites of both influenza A and B NAs with the *n*-butyl side chain adopting two different binding modes. In the B version, the side chain is positioned against a hydrophobic surface formed by Ala-246, Ile-222, and Arg-224. In the A version, the chain is in a region formed by the reorientation of the side chain of Glu-276.

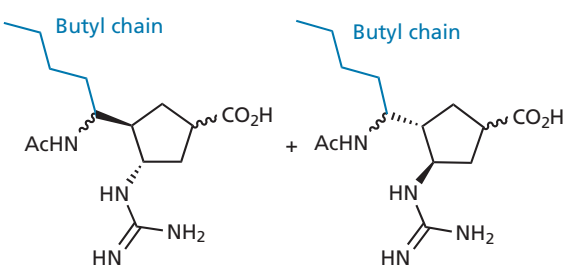


FIGURE 20.51 Mixture of isomers tested for their binding affinity to crystals of the neuraminidase enzyme.

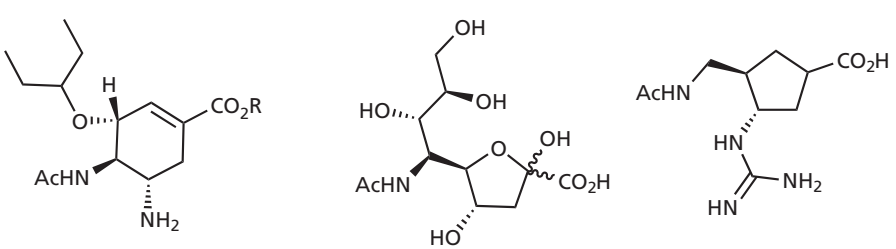


FIGURE 20.50 Oseltamivir and other ring systems.

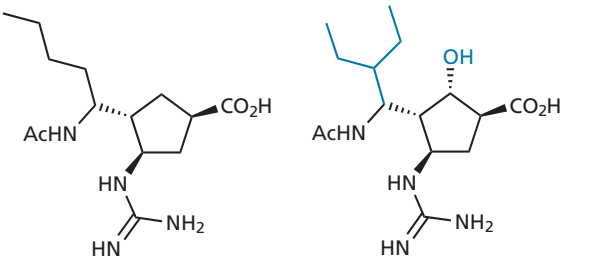


FIGURE 20.52 Development of peramivir (BCX 1812).

Peramivir (Fig. 20.52) was designed to take advantage of both hydrophobic pockets in the active site. It was prepared as a racemic mixture, and a crystal of the neuraminidase enzyme was used to bind the active isomer. Once identified, this was then prepared by a stereospecific synthesis. The relative stereochemistry of the substituents was the same as in structure (I).

In vitro tests of peramivir versus strains of influenza A and B show it to be as active as zanamivir and GS4071. It is also four orders of magnitude less active against bacterial and mammalian NAs, making it a potent and highly specific inhibitor of flu virus neuraminidase. *In vivo* tests carried out on mice showed it to be orally active and the compound was approved in Japan in 2010.

20.8.3.6 Resistance studies

Studies have been carried out to investigate the likelihood of viruses acquiring resistance to the drugs mentioned above. This is done by culturing the viruses in the presence of the antiviral agents to see if mutation leads to a resistant strain.

Zanamivir has a broad spectrum efficacy against all type A and B strains tested, and interacts only with conserved residues in the active site of NA. Thus, in order to gain resistance, one of these important amino acids has to mutate. A variant has been observed where Glu-119 has been mutated to glycine. This has reduced affinity for zanamivir and the virus can replicate in the presence of the drug. Removing Glu-119 affects the binding interactions with the 4-guanidinium group of zanamivir without affecting interactions with sialic acid. Zanamivir-resistant mutations were also found where a mutation occurred in HA around the sialic acid binding site. This mutation weakened affinity for sialic acid and so lowered binding. Thus, mutant viruses were able to escape more easily from the infected cell after budding. No such mutations have appeared during clinical trials, however.

Another mutation has been observed where Arg-292 is replaced by lysine. In wild-type NA, Arg-292 binds to the carboxylate group of the inhibitor and is partly responsible for distorting the pyranose ring from the chair to the boat conformation. In the mutant structure, the amino group of Lys-292 forms an ionic interaction with Glu-276 which normally binds the hydroxyl groups at positions 8 and 9 of the glycerol side chain. This results in a weaker interaction with inhibitors and substrate alike, leading to a weaker enzyme.

One conclusion that has been made from studies on easily mutatable targets is the desirability to find an inhibitor which is modified as little as possible from the normal substrate, and which uses the same interactions for binding.

KEY POINTS

- The flu virus contains (–) ssRNA and has two glycoproteins called haemagglutinin (HA) and neuraminidase (NA) in its outer membrane.
- HA binds to the sialic acid moiety of glycoconjugates on the outer surface of host cells leading to adsorption and cell uptake.
- NA catalyses the cleavage of sialic acid from glycoconjugates. It aids the movement of the virus through mucus and releases the virus from infected cells after budding.
- HA and NA act as antigens for flu vaccines. However, the influenza A virus readily mutates these proteins and new flu vaccines are required each year.
- The adamantanes are antiviral agents which inhibit influenza A by blocking a viral ion channel called the matrix (M2) protein. At high concentration they buffer the pH of endosomes. They are ineffective against influenza B which lacks the matrix (M2) protein.
- Neuraminidase has an active site which remains constant for the various types and subtypes of the enzyme, and which is different from the active sites of comparable mammalian enzymes.
- There are four important binding pockets in the active site. The sialic acid moiety is distorted from its normal chair conformation when it is bound.
- The mechanism of the enzyme-catalysed reaction is proposed to go through an endocyclic sialosyl cation transition state. Inhibitors were designed to mimic this state by introducing an endocyclic double bond.
- Successful antiviral agents have been developed using structure-based drug design.
- Different scaffolds can be used to hold the four important binding groups.
- There is an advantage in designing drugs which use the same binding interactions as the natural ligand when the target undergoes facile mutations.

20.9 Antiviral drugs acting against RNA viruses: cold virus

The agents used against flu are ineffective against colds, as these infections are caused by a different kind of virus called a rhinovirus. Colds are less serious than flus. Nevertheless, research has taken place to find drugs which can combat them.

There are at least 89 serotypes of **human rhinoviruses** (HRV) and they belong to a group of viruses called the **picornaviruses** which include the polio, hepatitis A, and foot and mouth disease viruses. They are among the smallest of the animal RNA viruses, containing a positive strand of RNA coated by an icosahedral shell made up of 60 copies of four distinct proteins, VP1–VP4 (Fig. 20.53). The proteins VP1–VP3 make up the surface of the virion. The smaller VP4 protein lies underneath to form the inner surface and is in contact with the viral RNA. At the junction between each VP1 and VP3 protein, there is a broad canyon 25 Å deep, and this is where attachment takes place between the virus and the host cell. On the canyon floor there is a pore which opens into a hydrophobic pocket within the VP1 protein. This pocket is either empty or occupied by a small molecule called a **pocket factor**. So far the identity of the pocket factor has not been determined but it is known from X-ray crystallographic studies that it is a fatty acid containing seven carbon atoms.

Test your understanding and practise your molecular modelling with Exercise 20.6.

When the virus becomes attached to the host cell, a receptor molecule on the host cell fits into the canyon and induces conformational changes that cause the VP4 protein and the *N*-terminus of VP1 to move to the exterior of the virus—a process called **externalization**. This is thought to be important to the process by which the virus is uncoated and releases its RNA into the host cell. It is thought that the pocket factor stabilizes the capsid when it is bound, and prevents the conformational changes that are needed to cause infection.

A variety of drugs having antiviral activity are thought to mimic the pocket factor by displacing it and binding to the same hydrophobic pocket. The drugs concerned are called **capsid-binding agents** and are characteristically long-chain hydrophobic molecules. Like the pocket factor, they stabilize the capsid by locking it into a stable conformation and preventing the conformational changes required for uncoating. They also raise the canyon floor and prevent the receptor on the host cell from fitting the canyon (Fig. 20.54).

Pleconaril (Fig. 20.55) is one such drug that has undergone phase III clinical trials which demonstrate

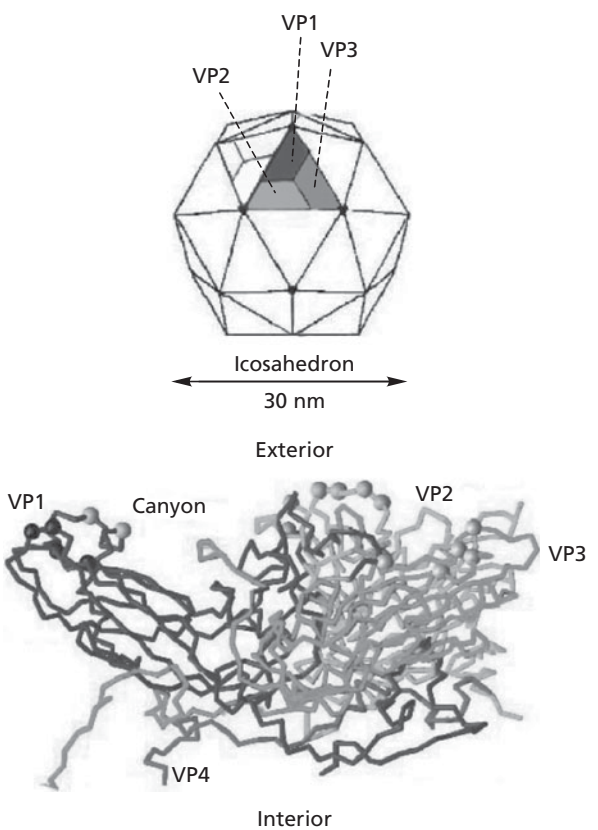


FIGURE 20.53 Structure of human rhinovirus and the proteins VP1–VP4.

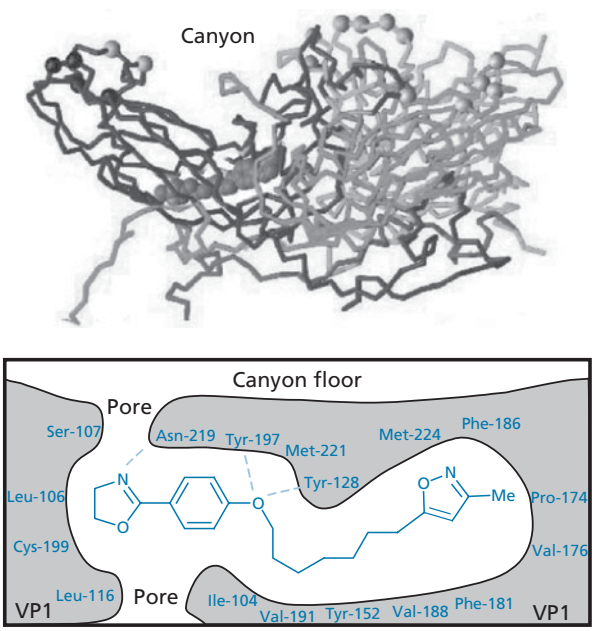


FIGURE 20.54 Binding of disoxaril (possible hydrogen bonds shown as dashed lines).

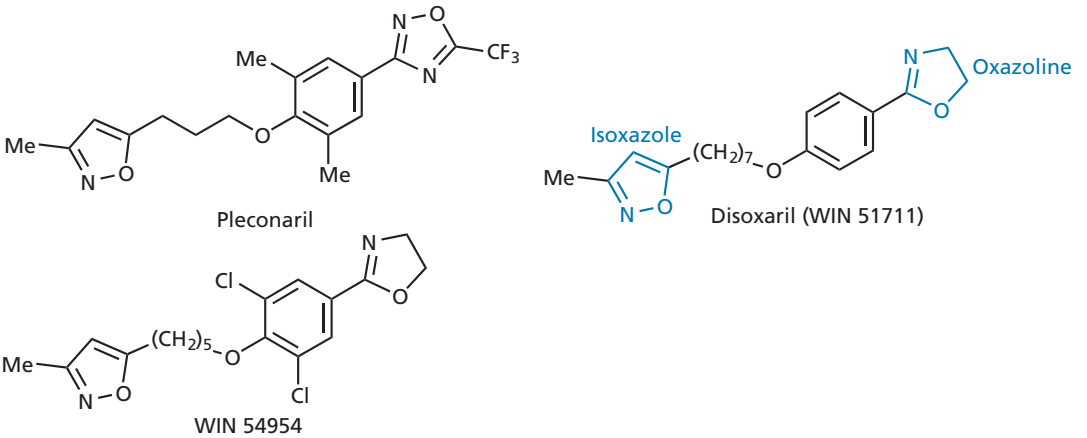


FIGURE 20.55 Capsid binding agents.

that it has an effect on the common cold. It is an orally active, broad-spectrum agent which can cross the blood–brain barrier. The drug may also be useful against the enteroviruses that cause diarrhoea, viral meningitis, conjunctivitis, and encephalitis, as these viruses are similar in structure to the rhinoviruses.

The development of pleconaril started when a series of isoxazoles were found to have antiviral activity. This led to the discovery of **disoxaril** (Fig. 20.55) which entered phase I clinical trials, but proved to be too toxic. X-ray crystallographic studies of VP1-drug complexes involving disoxaril and its analogues showed that the oxazoline and phenyl rings were roughly coplanar and were located in a hydrophilic region of the pocket near the pore leading into the centre of the virion (Fig. 20.54). The hydrophobic isoxazole ring binds into the heart of the hydrophobic pocket and the chain provides sufficient flexibility for the molecule to bend round a corner in the pocket. Ligand binding moves Met-221 which normally seals off the pocket. This also causes conformational changes in the canyon floor. Structure-based drug design was carried out to find safer and more effective antiviral agents. For example, the chain cannot be too short or too long, or else there are steric interactions. Placing additional hydrophobic groups on to the phenyl ring improves activity against the HRV2 strain, because increased interactions are possible with a phenylalanine residue at position 116 rather than leucine. The structure **WIN 54954** was developed and entered clinical trials, but results were disappointing because extensive metabolism resulted in 18 different metabolic products due mainly to hydrolysis of the oxazoline ring. Further structure-based drug design led to modifications of the phenyl and oxazoline moieties. This included the introduction of a trifluoromethyl group to block metabolism, resulting in pleconaril, with 70% oral bioavailability.

20.10 Antiviral drugs acting against RNA viruses: hepatitis C

Hepatitis C virus (HCV) is a positive-stranded RNA virus that was discovered in 1989. It is a blood-borne virus that affects an estimated 170 million people worldwide, but many of those infected with the agent are unaware of the fact as they do not experience any symptoms. However, the virus can cause serious liver damage, cancer, and, in the long term, death. Until recently, the only therapies available were the broad spectrum agents **pegylated alpha interferon** (**IFN-α**) and **ribavirin** (section 20.11). However, the success rates with these drugs are only about 40%. In May 2011, two new drugs with a more selective mode of action were approved for the treatment of hepatitis C—**boceprevir** and **telaprevir** (Fig. 20.56).

The life cycle of the virus within the host cell includes the synthesis of a 3000-amino acid polyprotein, which is cleaved into individual viral proteins by a viral protease enzyme called **HCV NS3-4A protease**. This is a serine protease containing a catalytic triad of Asp, His, and Ser (section 3.5.3). A study of the protein showed that the active site was a long shallow groove, and it was also found that the enzyme cleaved the peptide bond between cysteine and serine/alanine residues.

Having identified the target and the active site, it was decided to design an inhibitor that could interact with the active site, but would not undergo the enzyme-catalysed reaction. A series of peptide structures was studied where the susceptible amide bond was replaced with a keto amide group in the expectation that the serine residue in the active site would react with the ketone carbonyl group rather than the amide carbonyl group (Fig. 20.57). As a ketone group undergoes

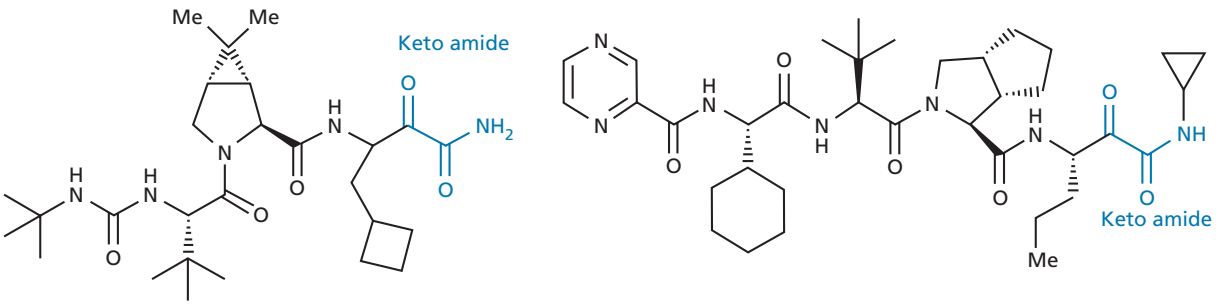


FIGURE 20.56 Boceprevir and telaprevir.

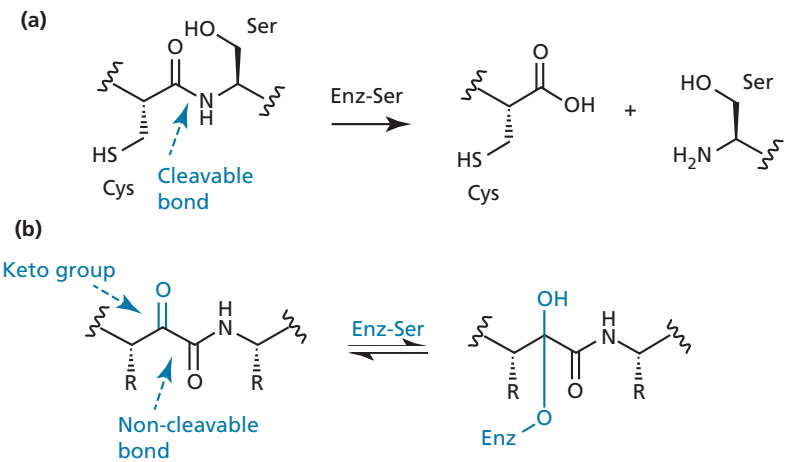


FIGURE 20.57 Design of protease inhibitors for HCV NS3-4A protease. (**a**) Normal enzyme-catalysed reaction. (**b**) Interaction of inhibitors with an active site serine residue.

nucleophilic addition rather then nucleophilic substitution, no bond cleavage results and a reversible covalent bond is formed between the inhibitor and the active site. A series of peptide structures containing the ketoamide group was screened, leading to the identification of an undecapeptide which reacted as planned and showed good activity as an inhibitor. This was adopted as the lead compound for further development; then, similar tactics to those used in the development of the HIV PIs were employed to design peptidomimetic agents that had increased binding interactions, decreased peptide character, and lower molecular weight. Boceprevir and telaprevir were the successful fruits of this labour. A second-generation family of these agents is currently undergoing clinical trials, for example **narlaprevir** (Fig. 20.58).

Research is also being carried out to develop selective agents that will inhibit an HCV enzyme called **NS5B RNA-dependent RNA polymerase**. One such agent undergoing clinical trials is **setrobuvir** (Fig. 20.59). Another target that is being studied is **non-structural protein 5A** (NS5A).

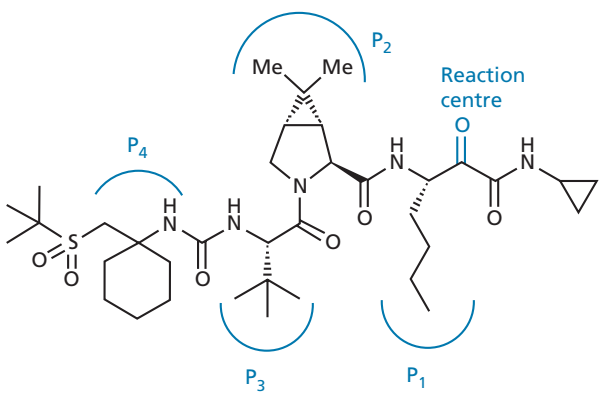


FIGURE 20.58 Narlaprevir (SCH 900518) with binding pockets indicated.

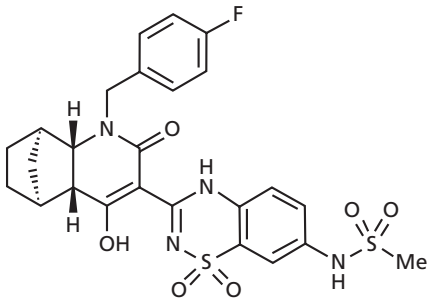


FIGURE 20.59 Setrobuvir.

20.11 Broad-spectrum antiviral agents

There are very few clinically useful, broad-spectrum antiviral agents that act on specific targets. The following are some examples.

20.11.1 Agents acting against cytidine triphosphate synthetase

Cytidine triphosphate is an important building block for RNA synthesis and so blocking its synthesis inhibits the synthesis of viral mRNA. The final stage in the biosynthesis of cytidine triphosphate is the amination of uridine triphosphate—a process that is catalysed by the enzyme cytidine triphosphate synthetase. **Cyclopentenyl cytosine** (Fig. 20.60) is a carbocyclic nucleoside that is converted in the cell to the triphosphate, which then inhibits this final enzyme in the biosynthetic pathway. The drug has broad antiviral activity against more than 20 RNA and DNA viruses, and has also been studied as an anticancer drug.

20.11.2 Agents acting against *S*-adenosylhomocysteine hydrolase

The 5′-end of a newly transcribed mRNA is capped with a methyl group in order to stabilize it against phosphatases and nucleases, as well as enhancing its translation. *S*-adenosylhomocysteine hydrolase is an intracellular enzyme that catalyses this reaction and many viruses need it to cap their own viral m-RNA. **3-Deazaneplanocin A** (Fig. 20.60) is an analogue of cyclopentenyl cytosine, and acts against a range of RNA and DNA viruses by inhibiting *S*-adenosylhomocysteine hydrolase.

20.11.3 Ribavirin

Ribavirin (Fig. 20.60) is a synthetic nucleoside that induces mutations in viral genes and is used against hepatitis C infection (section 20.10). It was the first synthetic, non-interferon-inducing broad-spectrum antiviral nucleoside and can inhibit both RNA and DNA viruses by a variety of mechanisms, although it is only licensed for hepatitis C and respiratory syncytial virus. Nevertheless, it has been used in developing countries for the treatment of tropical and haemorrhagic fevers, such as Lassa fever, when there is no alternative effective treatment. Tests show that it is useful in combination with other drugs such as rimantadine. Its dominant mechanism of action appears to be depletion of intracellular pools of GTP by inhibiting **inosine-5′-monophosphate dehydrogenase**. Phosphorylation of ribavirin results in a triphosphate which inhibits **guanyl transferase** and prevents the 5′ capping of mRNAs. The triphosphate can also inhibit **viral RNA-dependent RNA polymerase**. Owing to these multiple mechanisms of action, resistance is rare. The drug's main side effect is anaemia and it is a suspected **teratogen**.

20.11.4 Interferons

Interferons are small natural proteins that were discovered in 1957 and which are produced by host cells as a response to 'foreign invaders'. Once produced, interferons inhibit protein synthesis and other aspects of viral replication in infected cells. In other words, they shut the cell down. This can be described as an intracellular immune response. Administering interferons to patients

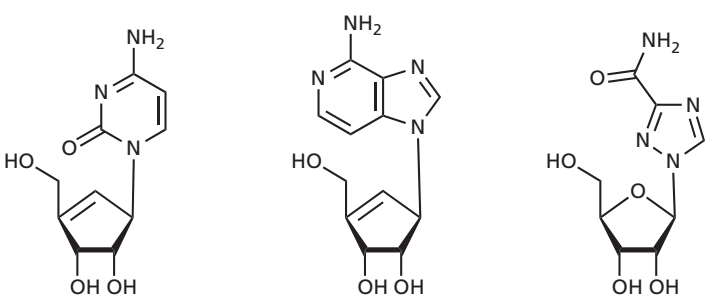


FIGURE 20.60 Broad-spectrum antiviral agents.

has been seen as a possible approach to treating flu, hepatitis, herpes, and colds.

There are several interferons which are named according to their source: α-interferons from lymphocytes, β-interferons from fibroblasts, and γ-interferons from T-cells. α-Interferon (also called **alferon** or **IFN-alpha**) is the most widely used of the three types. In the past, it was difficult and expensive to isolate interferons from their natural cells, but recombinant DNA techniques allow the production of genetically engineered interferons in larger quantities (section 6.4). Recombinant α-interferon is produced in three main forms. The α-2a and α-2b are natural forms, and **alfacon-1** is the unnatural form. They have proved successful therapeutically, but can have serious toxic side effects. At present, α-interferon is used clinically against hepatitis B infections. It is also used with ribavirin against hepatitis C infections.

Interferon production in the body can also be induced by agents known as **immunomodulators**. One such example is **avridine** (Fig. 20.61), which is used as a vaccine adjuvant for the treatment of animal diseases such as foot and mouth. **Imiquimod** (Fig. 20.61) also induces the production of α-interferon, as well as other cytokines that stimulate the immune system. It is effective against genital warts.

20.11.5 Antibodies and ribozymes

Antibodies that recognize a virion-specific antigen will bind to that antigen and mark the virus out for destruction by the body's immune system. **Palivizumab** is a humanized monoclonal antibody which was approved in 1998 for the treatment of respiratory syncytial infection in babies. It blocks viral spread from cell to cell by targeting a specific protein of the virus. Another monoclonal antibody is being tested for the treatment of hepatitis B.

It has been possible to identify sites in viral RNA that are susceptible to cutting by **ribozymes**—enzymatic forms of RNA. One such ribozyme is being tested in patients with hepatitis C and HIV. Ribozymes could be generated in the cell by introducing genes into infected cells—a form of gene therapy. Other gene therapy projects are looking at genes that would:

(i) code for specialized antibodies capable of seeking out targets inside infected cells, or
(ii) code for proteins that would latch on to viral gene sequences within the cell.

20.12 Bioterrorism and smallpox

Methisazone (Fig. 20.62) was the first effective antiviral drug to reach the clinic and was used to treat smallpox in the 1960s. The drug was no longer required once the disease was eradicated through worldwide vaccination. In recent years, however, there have been growing worries that terrorists might acquire smallpox and unleash it on a world no longer immunized against the disease. As a result, there has been a regeneration of research into finding novel antiviral agents which are effective against this disease.

KEY POINTS

- There are few broad-spectrum antiviral agents currently available.
- The best broad-spectrum antiviral agents appear to work on a variety of targets, reducing the chances of resistance.
- Interferons are chemicals produced in the body which shut down infected host cells and limit the spread of virus.
- Antibodies and ribozymes are under investigation as antiviral agents.

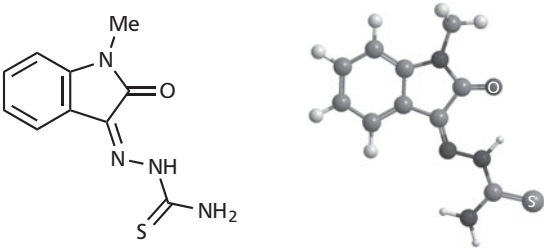


FIGURE 20.62 Methisazone together with ball and stick model.

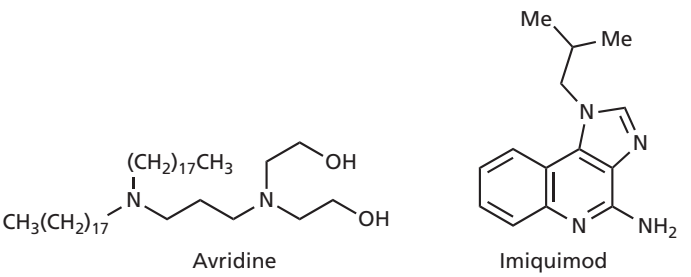
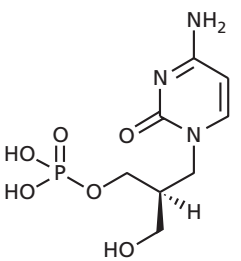


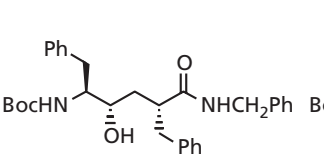
FIGURE 20.61 Immunomodulators.

QUESTIONS

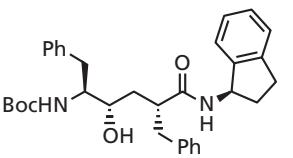
1. Consider the structures of the protease inhibitors (PIs) given in section 20.7.4 and suggest a hybrid structure that might also act as a PI.
2. Consider the structure of the PIs in section 20.7.4 and suggest a novel structure with an extended subsite ligand.
3. What disadvantage might the following structure have as an antiviral agent compared with cidofovir?



4. Zanamivir has a polar glycerol side chain which forms good hydrogen-bonding interactions with a binding pocket, yet carboxamides and oseltamivir have hydrophobic substituents which bind more strongly to this pocket. How is this possible?
5. Show the mechanism by which the prodrugs tenofovir disoproxil and adefovir dipivoxil are converted to their active forms. Why are extended esters used as prodrugs for these compounds?
6. Most PIs bind to the active site with a water molecule acting as a hydrogen bonding bridge to the enzyme flaps. Suggest what relevance this information might have in the design of novel PIs.
7. The following structures were synthesized during the development of L 685434 (Fig. 20.27). Identify the differences between the two structures and suggest why one is more active than the other.

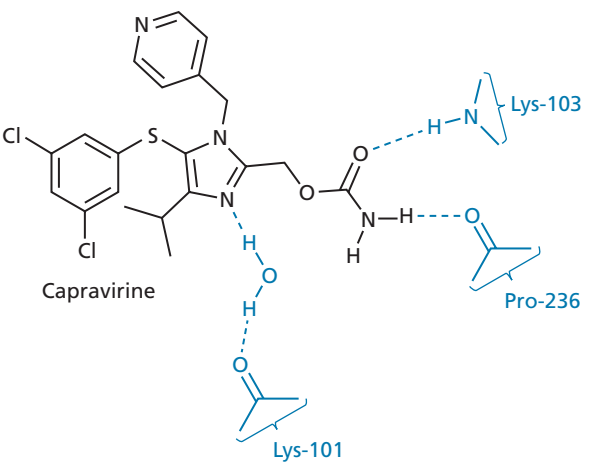


(IC_{50}= 111 nM)



(IC_{50}= 21 nM)

8. Capravirine is a third-generation non-nucleoside reverse transcriptase inhibitor (NNRTI) with a side chain that takes part in important hydrogen bonding to Lys-103 and Pro-236 in the allosteric binding site, yet the side chain has a carbonyl group. Discuss whether this makes the structure prone to enzymatic hydrolysis and inactivation.



FURTHER READING

Carr, A. (2003) Toxicity of antiretroviral therapy and implications for drug development. *Nature Reviews Drug Discovery* **2**, 624–634.

Coen, D. M. and Schaffer, P. A. (2003) Antiherpesvirus drugs: a promising spectrum of new drugs and drug targets. *Nature Reviews Drug Discovery* **2**, 278–288.

De Clercq, E. (2002) Strategies in the design of antiviral drugs. *Nature Reviews Drug Discovery* **1**, 13–25.

Driscoll, J. S. (2002) *Antiviral Agents*. Ashgate, Aldershot.

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.

Milroy, D. and Featherstone, J. (2002) Antiviral market overview. *Nature Reviews Drug Discovery* **1**, 11–12.

Tan, S-L., Pause, A., Shi, Y., and Sonenberg, N. (2002) Hepatitis C therapeutics: Current status and emerging strategies. *Nature Reviews Drug Discovery* **1**, 867–881.

Venkatraman, S., Bogen, S. L., Arasappan, A., Bennett, F., Chen, K., Jao, E., et al. (2006) Discovery of (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3-[2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]- 6,6-dimethyl-3-azabicyclo[3.1.0] hexan-2(S)-carboxamide (SCH 503034), a selective, potent, orally bioavailable hepatitis C virus NS3 protease inhibitor: a potential therapeutic agent for the treatment of

hepatitis C infection. *Journal of Medicinal Chemistry* **49**, 6074–6086 [boceprevir].

HIV

Beaulieu, P. R., Anderson, P. C., Cameron, D. R., Croteau, G., Gorys, V., and Grand-Maître, C., (2000) 2′,6′-Dimethylphenoxyacetyl: a new achiral high affinity P(3)-P(2) ligand for peptidomimetic-based HIV protease inhibitors. *Journal of Medicinal Chemistry* **43**, 1094–1108 [palinavir].

Campiani, G., Ramunno, A., Maga, G., Nacci, V., Fattorusso, C., Catalanotti, B., et al. (2002) Non-nucleoside HIV-1 reverse transcriptase (RT) inhibitors: past, present and future perspectives. *Current Pharmaceutical Design* **8**, 615–657.

De Clercq, E. (2003) The bicyclam AMD3100 story. *Nature Reviews Drug Discovery* **2**, 581–587.

Dubey, S., Satyanarayana, Y. D., and Lavinia, H. (2007) Development of integrase inhibitors for treatment of AIDS: An overview. *European Journal of Medicinal Chemistry* **42**, 1159–1168.

Ghosh, A. K. (2009) Harnessing nature's insight: Design of aspartyl protease inhibitors from treatment of drug-resistant HIV to Alzheimer's disease. *Journal of Medicinal Chemistry* **52**, 2163–2176 [darunavir and amprevanir].

Ghosh, A. K., Dawson, Z. L., and Mitsuya, H. (2007) Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV. *Bioorganic and Medicinal Chemistry* **15**, 7576–7580.

HIV databases. Available at: http://www.hiv.lanl.gov/content/index (last accessed April 2012).

Kempf, D. J. and Sham, H. L. (1996) Protease inhibitors. *Current Pharmaceutical Design* **2**, 225–246.

Matthews, T., Salgo, M., Greenberg, M., Chung, J., DeMasi, R., and Bolognesi, D. (2004) Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nature Reviews Drug Discovery* **3**, 215–225.

Miller, J. F., Furfine, E. S., Hanlon, M. H., Hazen, R. J., Ray, J. A., Robinson, L., et al. (2004) Novel arylsulfonamides possessing sub-picomolar HIV protease activities and potent anti-HIV activity against wild-type and drug-resistant viral strains. *Bioorganic and Medical Chemistry Letters* **14**, 959–963 [amprenavir].

Price, D. A., Armour, D., de Groot, M., Leishman, D., Napier, C., Perros, M., et al. (2006) Overcoming HERG affinity in the discovery of the CCR5 antagonist maraviroc. *Bioorganic and Medicinal Chemistry Letters* **16**, 4633–4637.

Raja, A., Lebbos, J., and Kirkpatrick, P. (2003) Atazanavir sulphate. *Nature Reviews Drug Discovery* **2**, 857–858.

Sansom, C. (2009) Molecules made to measure. *Chemistry World* November, 50–53.

Tomasselli, A. G. and Heinrikson, R. L. (2000) Targeting the HIV-protease in AIDS therapy. *Biochimica et Biophysica Acta* **1477**, 189–214.

Tomasselli, A. G., Thaisrivongs, S., and Heinrikson, R. L. (1996) Discovery and design of HIV protease inhibitors as drugs for AIDS. *Advances in Antiviral Drug Design* **2**, 173–228.

Werber, Y. (2003) HIV drug market. *Nature Reviews Drug Discovery* **2**, 513–514.

Flu

Babu, Y. S., Chand, P., Bantia, S., Kotian, P., Dehghani, A., El-Kattan, Y., et al. (2000) BCX-1812 (RWJ-270201): Discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *Journal of Medicinal Chemistry* **43**, 3482–3486.

Ezzell, C. (2001) Magic bullets fly again. *Scientific American* October, 28–35.

Kim, C. U., Lew, W., Williams, M. A., Wu, H., Zhang, L., Chen, X. et al. (1998) Structure–activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *Journal of Medicinal Chemistry* **41**, 2451–2460 [oseltamivir].

Laver, W. G., Bischofberger, N., and Webster, R. G. (1999) Disarming flu viruses. *Scientific American* January, 56–65.

Sansom, C. (2011) Fighting the flu. *Chemistry World* February, 44–47.

Taylor, N. R., Cleasby, A., Singh, O., Skarzynski, T., Wonacott, A. J., Smith, P. W., et al. (1998) Dihydropyrancarboxamides related to zanamivir: a new series of inhibitors of influenza virus sialidases. 2. crystallographic and molecular modeling study of complexes of 4-amino-4*H*-pyran-6-carboxamides and sialidase from influenza virus types A and B. *Journal of Medicinal Chemistry* **41**, 798–807 [carboxamides].

von Itzstein, M., Dyason, J. C., Oliver, S. W., White, H. F., Wu, W. Y., Kok, G. B., and Pegg, M. S. (1996) A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. *Journal of Medicinal Chemistry* **39**, 388–391 [zanamivir].

Titles for general further reading are listed on p. 763.

For additional material see Web article 22: case study on maraviroc, a CCR5 antagonist for HIV treatment