**Lecture V**

## **Prodrugs and their active metabolites**

**Prodrugs** are pharmacologically inactive derivatives of the active molecule that are designed to break down within the body to release the active drug. The prodrug approach is often used in pharmacy to overcome problems such as poor absorption or instability when the parent drug is given orally, or it the parent drug has an unpalatable taste or smell that needs to be disguised.



When a drug molecule gets converted into the body to an altogether different form, which may be either less or more active than the parent drug, the phenomenon is termed as **biotransformation**.

**Some prodrugs and its active metabolites.**

|  |  |
| --- | --- |
| Prodrugs |  Active metabolites |
|  1  |  2 |
| Benzobarbital (Benzonalum ) | Phenobarbitalum (Luminal) |
| Dipivefrine (Epinephrine-dipivalat) | Epinephrine (Adrenalinum) |
| Isosorbidedinitrate (Nitrosorbidum) | Isosorbidemononitrate |
| Enalapril (Enap,Enapril) | Enalarpilat |
| Azathioprinum (Imuran) | Mercaptopurine |
| Salazodine (salazopyridazinum) | Sulfapyridazinum |
| Salazodimethoxinum | Sulfadimethoxine |
| Calcium benzamidosalicylate (Bepascum) | Acidumparaaminosalicylicum |
| Methenamine (Urotropinum,Hexamethylentetraminum) | Formaldehydum |
| Phenyliisalicylas (Salolum) | Phenolum+Acidumsalicylicum |
| Phthorafurum (Tegafur) | Fluorouracil (Phthoruracilum) |
| Fosfestrol (Honvan) | Diethylstilbestrol |
| Primidone (Hexamidine) | Phenobarbitalum(LuminaL) |
| Diazepam (Sibazonum,Seduxen) | Oxazepam |
| Medazepam (Mezapamum) | Oxazepam |
| Picamilonum | Aminalonum+Nicotinic Acid |
| Levodapa (L-Dopa,Levopa) | Dophaminum+Norepinephrine |
| Codeinum | Morphine (Doltard) |
| Dophaminum (Dopamine) | Norepinephrine (Noradrenalinum) |
| Acetylcystein (Mucobene,Mucomyst) | Cysteinum |
| Phenylbutazone (Butadionum) | Sulfinpyrazone (Anturan) |
| Cortisone | Hydrocortisone |
| Estradiol (Proginova) | Estriol+Estrol |
| Thiamine (Vitamin B1) | Cocarboxylase(Pyruvodehydrase,Diphosphothiamin) |
| Monophosphothiamine(Phosphothiaminum) | Cocarboxylase(Pyruvodehydrase,Diphosphothiamin) |
| Riboflavinum (vitamin B2) | Riboflavinum-mononucleotidum+ Flavinatum |
| Pyridoxine (vitamin B6 ) | Pyridoxalphosphatum |
| Cyanocobalaminum (vitamin B12) | Cobamamide+Oxycobalaminum(Hydroxocobalamin) |
| Folic Acid (vitamin Bc) | Tetrahydrofolic Acid (Folin) |
| Nicotinic Acid (vitamin PP) | Nicotinamide |
| Alfacalcidol (Alfa D3) | Calcitriol |
| Riboxinum(Inosine,Inosie-F) | Acidumadenosintriphosphoricum(ATP) |
| Phthalylsulfathiazole(Phthalazolum) | Sulfathiazole (Norsulfazolum) |
| Phthalylsulfapyridazine(Phthazinum) | Sulfapyridazinum |
| Chinoxydinum | Dioxydinum |
| Solasulfone(Solusulfonum) | Dapsone (Diaphenylsulfonum) |
| Chloroquine(Chingaminum) | Hydroxychloroquine(PlaquieniL) |



 **Sulindac Active metabolic**

There are two main types of biotransformation observed in the body, imaginatively called **Phase 1**and **Phase 2** reactions, although many drugs undergo both types of process.

Phase 1 reactions are reactions in which a new functional group is introduced into the molecule, or an existing group is converted into another, (usually more water- soluble) derivative. **Functional group changes:** here, the “drug substance” undergoes functional group changes for instance : side-chain or ring hydroxylation, reduction of nitrogroup, reduction, aldehyde oxidation, deamination or dealkylation.

**Reductation**



 **Prontozyl Sulfonamide**

 **(inactive) (active)**

Phase 2 reactions (conjugations or esterification) are where an existing functional group in the molecule is masked by the addition of a new group.The conjugate is formed between the drug and a hydrophilic compound such as glucuronic acid and the resulting conjugate (a glucuronide ) will usually be much more water soluble than the parent drug. Most drugs are hydrophobic and so not inherently water soluble. Metabolism to a more water–soluble and less toxic derivative terminates drug action and allows the body to excrete the drug easily in the urine. If the administered drug is already hydrophilic, the molecule is often excreted unchanged.

Excretion of drugs from their sites of action is of paramount importance and may be effectively carried out with the help of a number of processes, namely:renal excretion, biliary excretion, excretion through lungs and above all by drug metabolism (biotransformation).

**Some prodrugs and medicine substances, formed from them**

|  |  |
| --- | --- |
| **Prodrug** | **Drug** |
| Azatioprin | 6-Merkaptopurin |
| Balsalazid | Mesalazin |
| Xloramfenikol-suksinat | Xloramfenikol |
| Dipiverine | Epinefrin |
| Enalapril | Enalaprilat |
| Haloperidol-dekanoat | Haloperidol |
| Olsalazin | Mesalazin |
| Pivampisillin | Ampisillin |
| Proquanil (Biqumal) | Tsikloquanil |
| Propasetamol | Parasetamol |

Drugs which are either water-soluble or get metabolized gradually are mostly eliminated through the kidneys by the aid of these three essential phenomena,viz: secretion,glomerular filtration and tubular reabsorption. For instance, probenecid considerably retards tubular secretion of penicillin thereby enhancing its duration of action appreciably.

Another aspect of excretion is the biliary excretion of drugs or its metabolites which essentially affects excretion of drugs by liver cells into the bile subsequently into intestine. Invariably, a drug undergoes “enterohepatic cycling” ,i.e., instead of its elimination through the faeces it gains entry into the system through the intestines,eg.,penicillin,fluorescein,etc.

The processers involved in drug metabolism involve simple chemical reactions such as oxidation ( the most common) reduction and dealkylation and are influenced by a number of factors including:

**Prodrugs and their active metabolite**

|  |  |
| --- | --- |
| **Prodrug their İNN** **and trade names** | **Active metabolite** |
| **1** | **2** |
| Benzobarbital (Benzonal) | Fenobarbital (Lüminal) |
| Dipivefrin (Adrenalin - dipivalat) | Epinefrin (Adrenalin) |
| İzosorbid-dinitrat (Nitrosorbid) | İzosorbid- mononitrat |
| Enalapril (Enap, Enapril) | Enalaprilat |
| Azatioprin (İmuran) | Merkaptopurin |
| Salazodin (Salazopiridazin) | Sulfapiridazin |
| Kalsium-benzamidosalisilat (Bepask) | Past-(p –Aminsalisil turşusu) |
| Metenamin (Urotropin, Heksametilentetramin) | Formaldehid |
| Fenilsalisilat (Salol) | Fenol + Salisil turşusu |
| Ftorafur (Teqafur) | Flüorurasil |
| Fosfestrol (Honvan) | Dietilstilbestrol |
| Primidon (Heksamidin) | Fenobarbital |
| Diazepam (Sibazon, Seduksen) | Oksazepam |
| Medazepam (Mezepam) | Oksazepam |
| Pikamilon | Aminolon+Nikotin turşusu |
| Levodopa (L-Dofa, L-Dopa) | Dofamin + Noradrenalin |
| Kodein | Morfin (Doltard) |
| Dopamin (Dofamin) | Norepinefrin (Noradrenalin) |
| Asetilsistein (Mukobene, Nukomist) | Sistein |
| Fenilbutazon (Butadion) | Sulfinpirazon (Anturan) |
| Kortizon | Hidrokortizon |
| Estradiol (Proqinova) | Estron + Estriol |
| Tiamin (B1 vitamini) | Kokarboksilaza (Tiaminpirofosfat) |
| Monofosfotiamin (Fosfotiamin) | Kokarboksilaza (Tiaminpirofosfat) |
| 1 | 2 |
| Riboflavin (B2 vitamini) | Riboflavin mononukleotid +Flavinat |
| Piridoksin (B6 vitamini) | Piridoksalfosfat |
| Sianokobalamin (B12 vitamini) | Kobamamid + Oksikobalamin |
| Fol turşusu (Bc vitamini) | Tetrahidrofol (Folin) turşusu |
| Nikotin turşusu (PP vitamini) | Nikotinamid |
| Alfakalsidol (Alfa D2) | Kalsitriol |
| Riboksin (İnozie-F) | ATF |
| Ftalilsulfatiazol (Ftalazol) | Sulfatiazol (Norsulfazol) |
| Ftalilsulfapiridazin (Ftazin) | Sulfapiridazin |
| Xinoksidin | Dioksidin |
| Solasulfon (Solusulfon) | Dapson (Diafenilsulfon) |
| Xloroxin (Xinqamin) | Hidroksixloroxin (Plakvenil) |

**Genetic factors:** Differences are observed between species (important since most medicines intended for human use are tested first in animals) and between individuals in a population.

**Physiological factors:** These includeage ofthe patient,gender,pregnancy and nutritional status. Very young patients whose livers have not developed fully and very old patients whose liver function has deteriorated metabolise drugs more slowly than the normal adult population. There are also differences in the rates of metabolism between men and women and between pregnant and non-pregnant women.The causes of these effects are unknown but are probably due to differences in levels of circulating sex hormones.

**Pharmacodynamic factors:** Include dose, frequency and route of administration and extent of protein binding.

**Environmental factors:** Examples of these are co-administration of other drugs, which can effect the rate and extent of drug metabolism. This can become literally a matter of life and death as a number of potentially fatal drug interactions involve liver enzyme induction and competition for drug-metabolism enzymes.

 ;  ; ; ;;

 (1) (2) (3) (4) (5)

 ; ; ;  ;  ;

 (6) (7) (8) (9) (10)

 ; ; ;;  ;

 (11) (12) (13) (14) (15)

; ;  ;  ; 

 (16) (17) (18) (19) (20)

 **Scheme 1.**

**Metabolic pathways of drugs in the body**

Compounds with similar structures to a pharmacologically active drug are often themselves biologically active. This activity may be either similar to that of the original compound but different in potency and unwanted side effects or com- pletely different to that exhibited by the original compound. These structurally related activities are commonly referred to as structure–activity relationships (SARS). A study of the structure–activity relationships of a lead compound and its analogues may be used to determine the parts of the structure of the lead compound that are responsible for both its beneficial biological activity, that is, its pharmacophore, and also its unwanted side effects. This information may be used to develop a new drug that has increased activity, a different activity from an existing drug and fewer unwanted side effects.

 

*R= –CH3 (İmipramin, İmizin) R= –COOH (L-do­fa)*

*R= –H (Dezipramin) R= H (Dofamin)*



|  |  |
| --- | --- |
|  Еналаприл → | Еналаприлат |



 

 *Pivampisillin Bakampisillin Talampisillin*

Structure–activity relationships are usually determined by making minor changes to the structure of a lead to produce analogues (see section 2.3) and assessing the effect these structural changes have on biological activity. The investigation of numerous lead compounds and their analogues has made it possible to make some broad generalizations about the biological effects of specific types of structural change. These changes may be conveniently classified as changing

1. the size and shape of the carbon skeleton (see section 4.2),
2. the nature and degree of substitution (see section 4.3), and
3. the stereochemistry of the lead (see section 3.2).

50 2000

(CH2)nH x

HO

OH

x

x

x

x

xx

xxx

40

Antibacterial activity

30

IC50(nM)

n = 2, IC50 = 19,000

x

(CH2)n

1000

20

10

HOOC

N C N

H O

COOH

0

2 4 6 8 10

0 (19) x

2 3 4

(4.8)

x

5

x (8.1)

6

Increasing values of n Increasing values of n

1. **(b)**

Examples of the variation of response curves with increasing numbers of inserted methylene groups. (a) A study by Dohme *et al*. on the variation of antibacterial activity of 4-alkyl substituted resorcinols. (b) Inhibition of ACE by enalaprilat analogues (Thorsett). The figures in brackets are the IC50 values for that analogue

The selection of the changes required to produce analogues of a particular lead is made by considering the activities of compounds with similar structures and also the possible chemistry and biochemistry of the intended analogue. It is believed that structural changes that result in analogues with increased lipid character may exhibit either increased activity because of better membrane penetration (Figure 4.1(a); *n* ¼ 3–6) or reduced activity because of a reduction in their water solubility (Figure 4.1(b)). However, whatever the change, its effect on water solubility, transport through membranes, receptor binding, and metabolism and other pharmacokinetic properties of the analogue should be considered as far as is possible before embarking on what could be an expensive synthesis. Furthermore, changing the structure of the lead com- pound could result in an analogue that is too big to fit its intended target site. Computer assisted molecular modelling (see Chapter 5) can alleviate this problem, provided that the structure of the target is known or can be simulated with some degree of accuracy. However, it is emphasized that although it is possible to predict the effect of structural changes there will be numerous exceptions to the predictions, and so all analogues must be synthesized and tested.

Compounds with similar structures to a pharmacologically active drug are often themselves biologically active. This activity may be either similar to that of the original compound but different in potency and unwanted side effects or com- pletely different to that exhibited by the original compound. These structurally related activities are commonly referred to as structure–activity relationships (SARS). A study of the structure–activity relationships of a lead compound and its analogues may be used to determine the parts of the structure of the lead compound that are responsible for both its beneficial biological activity, that is, its pharmacophore, and also its unwanted side effects. This information may be used to develop a new drug that has increased activity, a different activity from an existing drug and fewer unwanted side effects.

Structure–activity relationships are usually determined by making minor changes to the structure of a lead to produce analogues (see section 2.3) and assessing the effect these structural changes have on biological activity. The investigation of numerous lead compounds and their analogues has made it possible to make some broad generalizations about the biological effects of specific types of structural change. These changes may be conveniently classified as changing

1. the size and shape of the carbon skeleton (see section 4.2),
2. the nature and degree of substitution (see section 4.3), and
3. the stereochemistry of the lead (see section 3.2).

50 2000

(CH2)nH x

HO

OH

x

x

x

x

xx

xxx

40

Antibacterial activity

30

IC50(nM)

n = 2, IC50 = 19,000

x

(CH2)n

1000

20

10

HOOC

N C N

H O

COOH

0

2 4 6 8 10

0 (19) x

2 3 4

(4.8)

x

5

x (8.1)

6

Increasing values of n Increasing values of n

1. **(b)**

Examples of the variation of response curves with increasing numbers of inserted methylene groups. (a) A study by Dohme *et al*. on the variation of antibacterial activity of 4-alkyl substituted resorcinols. (b) Inhibition of ACE by enalaprilat analogues (Thorsett). The figures in brackets are the IC50 values for that analogue

The selection of the changes required to produce analogues of a particular lead is made by considering the activities of compounds with similar structures and also the possible chemistry and biochemistry of the intended analogue. It is believed that structural changes that result in analogues with increased lipid character may exhibit either increased activity because of better membrane penetration (Figure 4.1(a); *n* ¼ 3–6) or reduced activity because of a reduction in their water solubility (Figure 4.1(b)). However, whatever the change, its effect on water solubility, transport through membranes, receptor binding, and metabolism and other pharmacokinetic properties of the analogue should be considered as far as is possible before embarking on what could be an expensive synthesis. Furthermore, changing the structure of the lead com- pound could result in an analogue that is too big to fit its intended target site. Computer assisted molecular modelling (see Chapter 5) can alleviate this problem, provided that the structure of the target is known or can be simulated with some degree of accuracy. However, it is emphasized that although it is possible to predict the effect of structural changes there will be numerous exceptions to the predictions, and so all analogues must be synthesized and tested.

**Changing size and shape**

The shapes and sizes of molecules can be modified in a variety of ways, such as changing the number of methylene groups in chains and rings, increasing or decreasing the degree of unsaturation and introducing or removing a ring system. These types of structural change usually result in analogues that exhibit either a different potency or a different type of activity to the lead.

###  Introduction of new substituents

The new substituents may either occupy a previously unsubstituted position in the lead compound (see section 4.3.1) or replace an existing substituent (see section 4.3.2). Each new substituent will impart its own characteristic chemical, pharmacokinetic and pharmacodynamic properties to the analogue. Over the years, a great deal of information has been collected about the changes caused to these properties of a lead compound when a new substituent is incorporated into its structure. As a result, it is possible to generalize about some of the changes caused by the introduction of a particular group into a structure (see Table 4.2). However, the choice of substituent will ultimately depend on the properties that the development team decide to enhance in an attempt to meet their objectives. Moreover, it should be realized that the practical results of such a structural change will often be different from the theoretical predictions.

####  The introduction of a group in an unsubstituted position

The incorporation of any group will always result in analogues with a different size and shape to the lead compound. In addition, it may introduce a chiral centre, which will result in the formation of stereoisomers, which may or may not have different pharmacological activities (Table 2.1). Alternatively, it may impose conformation restrictions on some of the bonds in the analogue.

The introduction of a new group may result in an increased rate of metabol- ism, a reduction in the rate of metabolism or an alternative route for metabolism. These changes could also change the duration of action and the nature of any side effects. For example, mono- and diortho-methylation with respect to the phenolic hydroxy group of paracetamol produces analogues with

Examples of the ways in which the size and shape of the carbon skeletons of lead compounds may be changed to produce new analogues

Change Notes Example, the lead compound is given in square brackets([ ])

The number of methylene (CH2) groups in a chain or ring.

The degree of unsaturation.

Increasing the number of CH2 groups in a *chain* can lead to micelle formation which can reduce drug activity (Fig. 4.1). Changing

the number of CH2 groups in a *ring* may lead to a change in activity.

Introduction of a double bond increases the rigidity of the structure and in some cases the possibility of E and Z isomers. The reduction of double bonds makes the structure more

S

N Cl

CH2CH2CH2N(CH3)2

Chlorpromazine (Antipsychotic)

CH2OH O OH

C

HO

O

N Cl

CH2CH2CH2N(CH3)2

Clomipramine (Antidepressant)

CH2OH

O OH

C

HO

O

flexible. Cortisol

(Anti-inflammatory)

Prednisone (Potency 30) *Note*. No E isomer is possible in

this example.

Addition or removal of a ring.

Introduction of a ring may result in the filling of a hydrophobic pocket in the target, which might improve the

CH3O CH3O

CH3O

O

O 3-(3,4-Dimethoxy phenyl)-butyrolactam

N

H (Antidepressant)

O

Rolipram

binding of the drug to its target.

The incorporation of

CH3 S

NHCOCH

N H (Potency 10)

Benzylpenicillin

larger ring systems may be used to produce analogues that are resistant to enzymic attack.

CH3 HOOC

CH3 CH3 HOOC

2

N

O

S NHCO

N

O

(not b-lactamase resistant)

Diphenicillin (b-lactamase resistant)

Removal of ring systems has been used to produce

analogues of naturally occuring active compounds.

OH CH3

CH N

3

O

OH

N O

C

OC2H5

Morphine (Narcotic analgesic)

Pethidine (Narcotic analgesic)

Examples of some of the groups commonly used as new substituents in the production of analogues

Group Effect on lipophilic character

Likely change in solubility (see sections 3.3 and

3.4)

Notes

Methyl Increased lipophilic character

Decreased water solubility.

Increased lipid solubility.

Improves ease of absorption but makes its release from biological membranes more difficult. Can lead to changes in the nature and rate of metabolism. Larger alkyl groups will have similar effects.

Fluorine and chlorine

Increased lipophilic character

Decreased water solubility.

Increased lipid solubility.

Used to improve ease of penetration of cell membranes. However, there is an undesireable tendency for halogenated drugs to accumulate in lipid tissues. CF3 groups are sometimes used to replace Cl groups as these groups are of a similar size.

Hydroxy Decreased lipophilic character

Amino groups Decreased lipophilic character

Increased water solubility.

Decreased lipid solubility.

Increased water solubility due to salt formation. Decreased lipid solubility.

Provides a new centre for hydrogen, which could influence the binding of the drug to the target site. The presence of the hydroxy group could result in an increase in the rate of elimination of the drug by a new metabolic pathway and/or excretion.

Provides a new centre for hydrogen bonding, which could influence the binding of the drug to the target site. The incorporation of aromatic amines is avoided as they are often toxic and/or carcinogenic.

Carboxylic and sulphonic groups

Decreased lipophilic character

Increased water solubility due to salt formation. Decreased lipid solubility

Water solubility may be enhanced by *in vivo* salt formation.

Introduction usually increases the ease of elimination. Carboxylic acid group introduction into small lead molecules may change the type of activity of the analogue whilst sulphonic acid group incorporation does not

normally change the type of activity.

Steric hindrance between the hydrogen atom and the lone pairs.

O

H H C

H**..**

O**..**

N N

Diphenhydramine o-Methyl analogue

Harmes *et al*. suggest that the lack of antihistamine activity in the ortho-methyl analogue of diphenyhydramine is due to the ortho-methyl group restricting rotation about the C–O bond. It is believed that this prevents the molecule from adopting the conformation necessary for antihistamine activity reduced hepatotoxicity. It is believed that this reduction is due to the methyl groups preventing metabolic hydroxylation of these ortho positions.

HO NHCOCH3

CH3 HO

NHCOCH3

Paracetamol

CH3 o,o'-Dimethyl analogue of paracetamol

The position of substitution is critical. In one position the new group will lead to an enhancement of activity, while in another position it will result in a reduction of activity. For example, the antihypertensive clonidine with its o,o’-dichloro substitution is more potent than its m,p-dichloro analogue.

####  The introduction of a group by replacing an existing group

Analogues formed by replacing an existing group by a new group may exhibit the general stereochemical and metabolic changes outlined in section 4.3.1. The choice of group will depend on the objectives of the design team. It is often made using the concept of *isosteres*. Isosteres are groups that exhibit some similarities

Cl HN

Cl

Cl

HN

NH

NH

N Cl N

Clonidine ED 0.01 mg kg1 ED 3.00 mg kg1

20 20

Clonidine and its m,p-dichloro analogue. It is believed that the bulky chloro groups impose a conformation restriction on clonidine, which probably accounts for its greater activity Examples of isosteres. Each horizontal row represents a group of structures that are isosteric. Classical isosteres were originally defined by Erlenmeyer as atoms, ions and molecules with identical shells of electrons. Bioisosteres are groups with similar structures that usually exhibit similar biological activities

###  Quantitative structure–activity relationships (QSARS)

QSAR is an attempt to remove the element of luck from drug design by establishing a mathematical relationship in the form of an equation between biological activity and measurable physicochemical parameters. These param- eters are used to represent properties such as lipophilicity, shape and electron distribution, which are believed to have a major influence on the drug’s activity. They are normally defined so that they are in the form of numbers, which are derived from practical data that is thought to be related to the property the parameter represents. This makes it possible to either to measure or to calculate these parameters for a group of compounds and relate their values to the biological activity of these compounds by means of mathematical equations using statistical methods such as regression analysis (see Appendix 6). These equations may be used by the medicinal chemist to make a more informed choice as to which analogues to prepare. For example, it is often possible to use statistical data from other compounds to calculate the theoretical value of a specific parameter for an as yet unsynthesized compound. Substituting this value in the appropriate equation relating activity to that parameter, it is possible to calculate the theoretical activity of this unknown compound. Alter- natively, the equation could be used to determine the value *‘x’* of the parameter *‘y’* that would give optimum activity. As a result, only analogues that have values of *y* in the region of *x* need be synthesized.

The main properties of a drug that appear to influence its activity are its, lipophilicity, the electronic effects within the molecule and the size and shape of the molecule (steric effects). Lipophilicity is a measure of a drug’s solubility in lipid membranes. This is usually an important factor in determining how easily a drug passes through lipid membranes (see Appendix 5). The electronic effects of the groups within the molecule will affect its electron distribution, which in turn has a direct bearing on how easily and permanently the molecule binds to its target molecule (see Chapter 7). Drug size and shape will determine whether the drug molecule is able to get close enought to its target site in order to bind to that site. The parameters commonly used to represent these properties are partition coeffi- cients for lipohilicity (see section 4.4.1), Hammett s constants for electronic effects (see section 4.4.2) and Taft *M*s steric constants for steric effects (see section 4.4.3). Consequently, this text will be largely restricted to a discussion of the use of these constants. However, the other parameters mentioned in this and other texts are normally used in a similar fashion.

QSAR derived equations take the general form: biological activity ¼ function{parameter(s)}in which the activity is normally expressed as log[1/(concentration term)], usu- ally *C,* the minimum concentration required to cause a defined biological response. Where there is a poor correlation between the values of a specific parameter and the drug’s activity, other parameters must be playing a more important part in the drug’s action, and so they must also be incorporated into the QSAR equation.

QSAR studies are normally carried out on groups of related compounds. However, QSAR studies on structurally diverse sets of compounds are becom- ing more common. In both instances it is important to consider as wide a range of parameters as possible.

####  Lipophilicity

Two parameters are commonly used to represent lipophilicity, namely the partition coefficient (*P*) and the lipophilicity substituent constant (p). The former parameter refers to the whole molecule whilst the latter is related to substituent groups.

##### Partition coefficients (P)

A drug has to pass through a number of biological membranes in order to reach its site of action. Consequently, organic medium/aqueous system partition coeffi- cients were the obvious parameters to use as a measure of the ease of movement of the drug through these membranes. The accuracy of the correlation of drug activity with partition coefficients will depend on the solvent system used as a model for the membrane. A variety of organic solvents, such as n-octanol, chloroform and olive oil, are used to represent the membrane (organic medium), whilst both pure water and buffered solutions are used for the aqueous medium. The n-octanol–water system is frequently chosen because it appears to be a good mimic of lipid polarity and has an extensive database. However, more accurate results may be obtained if the organic phase is matched to the area of biological activity being studied. For example, n-octanol usually gives the most consistent results for drugs absorbed in the GI tract whilst less polar solvents such as olive oil frequently give more consistent correlations for drugs crossing the blood–brain barrier. More polar solvents such as chloroform give more consistent values for buccal absorption (soft tissues in the mouth).

The nature of the relationship between *P* and drug activity depends on the range of *P* values obtained for the compounds used. If this range is small the results may be expressed as a straight line equation having the general form:

where *k*1 and *k*2 are constants. This equation indicates a linear relationship between the activity of the drug and its partition coefficient. Over larger ranges of *P* values the graph of log 1/*C* against log *P* often has a parabolic form (Figure 4.5) with a maximum value (log *P*0). The existence of this maximum value implies that there is an optimum balance between aqueous and lipid solubility for maximum biological activity. Below *P*0 the drug will be reluctant to enter the membrane whilst above *P*0 the drug will be reluctant to leave the membrane. Log *P*0 represents the optimum partition coefficient for biological activity. This means that analogues with partition coefficients near this optimum value are likely to be the most active and worth further investigation. Hansch *et al*. showed that many of these parabolic relationships could be represented reason- ably accurately by equations of the form:

log (1=*C* ) ¼ *k*1( log *P*)2 þ *k*2 log *P* þ *k*3 (4:3)

where *k*1, *k*2 and *k*3 are constants that are normally determined by regression analysis.

log (1/*C* )

log *P* 0 log *P*

Figure 4.5 A parabolic plot for log (1/*C* ) against log *P*

##### Lipophilic substituent constants (p)

Lipophilic substituent constants are also known as hydrophobic substituent constants. They represent the contribution that a group makes to the partition coefficient and were defined by Hansch and co-workers by the equation:

p ¼ log *P*RH log *P*RX (4:4)

where *P*RH and *P*RX are the partition coefficients of the standard compound and its monosubstituted derivative respectively. However, when several substituents are present, the value of p for the compound is the sum of the p values of each of the separate substituents.

The value of p for a specific substituent will vary with the structural environ- ment of the substituent. Consequently, average values or the values relevant to the type of structure being investigated may be used in determining activity relationships. It also depends on the solvent system used to determine the partition coefficients. The values of p will also depend on the solvent system used to determine the partition coefficients used in their calculation. Most values are determined using the n-octanol/water system. A positive p value indicates that a substituent has a higher lipophilicity than hydrogen and so will probably increase the concentration of the com- pound in the n-octanol layer and by inference its concentration in the lipid material of biological systems. Conversely, a negative p value shows that the substituent has a lower lipophilicity than hydrogen and so probably increases the concentration of the compound in the aqueous media of biological systems. Furthermore, biological activity–p relationships that have high regression constants (Appendix 6) and low standard deviations demonstrate that the substituents are important in determining the lipophilic character of the drug.

Lipophilic constants are frequently used when dealing with a series of ana- logues in which only the substituents are different. This usage is based on the assumption that the lipophilic effect of the unchanged part of the structure is similar for each of the analogues.Electronic effects

The distribution of the electrons in a drug molecule has a considerable influence on the distribution and activity of a drug. In general, nonpolar and polar drugs in their unionized form are more readily transported through membranes than polar drugs and drugs in their ionized forms. Furthermore, once the drug reaches its target site the distribution of electrons in its structure will control the type of bond it forms with that target, which in turn affects its biological activity. The first attempt to quantify the electronic affects of groups on the physicochemical properties of compounds was made by Hammett (ca. 1940).

##### The Hammett constant (s)

The distribution of electrons within a molecule depends on the nature of the electron withdrawing and donating groups found in that structure. Hammett used this concept to calculate what are now known as Hammett constants (sX ) for a variety of monosubstituted benzoic acids (Equation (4.5) ). He used these constants to calculate equilibrium and rate constants for chemical reactions. However, they are now used as electronic parameters in QSAR relationships. Hammett constants (sX) are defined as:

where *K*B and *K*BX are the equilibrium constants for benzoic acid and mono- substituted benzoic acids respectively. Its value varies depending on whether the substituent is an overall electron donor or acceptor. A negative value for sX indicates that the substituent is acting as an electron donor group since *K*B > *K*BX. Conversely, a positive value for sX shows that the substituent is acting as an electron withdrawing group as *K*B < *K*BX. The value of sX for a specific substituent contains both inductive and mesomeric (resonance) contributions, and so varies with the position of that substituent in the molecule. This variation is indicated by the use of the subscripts m and p (Table 4.5). Inductive and Swain–Lupton constants are attempts to quantify the inductive and mesomeric effects of a substituent.

Hammett postulated that the s values calculated for the ring substituents of a series of benzoic acids could also be valid for those ring substituents in a different series of similar aromatic compounds. This relationship has been found to be in good agreement for the meta and para substituents of a wide variety of aromatic compounds but not for their ortho substituents. The latter is believed to be due to steric hindrance and other effects, such as intramolecular hydrogen bonding.

Hammett substitution constants suffer from the disadvantage that they only apply to substituents directly attached to a benzene ring. Consequently, a number of other electronic constants (Table 4.5) have been introduced and used in QSAR studies in a similar manner to the Hammett constants. However, attempts to relate biological activity exclusively to the values of Hammett substitution and similar constants have been largely unsuccessful, since electron distribution is not the only factor involved (see section 4.4).

Steric effects

The first parameter used to show the relationship between the shape and size (bulk) of a drug, the dimensions of its target site and the drug’s activity was

the Taft steric parameter (*E*s). It was followed by Charton’s steric parameter (n), Verloop’s steric parameters and the molar refractivity (MR) amongst others. The most used of these additional parameters is probably the molar refractivity.

#####  The Taft steric parameter (Es)

Taft (1956) used the relative rate constants of the acid catalysed hydrolysis of a-substituted methyl ethanoates to define his steric parameter because it had been shown that the rates of these hydrolyses were almost entirely dependent on steric factors. He used methyl ethanoate as his standard and defined *E*s as:

Taft steric parameters have been found to be useful in a number of investi- gations (see section 4.4.4). They also suffer from the disadvantage that they are determined by experiment. This has limited the number of values recorded in the literature.

#####  Molar refractivity ( MR)

The molar refractivity is a measure of both the volume of a compound and how easily it is polarized. It is defined as:

where *n* is the refractive index, *M* the relative mass and r the density of the compound. The *M*/r term is a measure of the molar volume whilst the refractive index term is a measure of the polarizability of the compound. Although MR is calculated for the whole molecule, it is an additive parameter, and so the MR values for a molecule can be calculated by adding together the MR values for its component parts.

Examples of calculated MR values. Reproduced by permission of John Wiley and Sons Ltd. from Hansch C. and Leo A.J. *Substituents Constants for Correlation Analysis in Chemistry and Biology* (1979)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Group | MR | Group | MR | Group | MR |
| H– | 1.03 | F– | 0.92 | CH3O– | 7.87 |
| CH3– | 5.65 | Cl– | 6.03 | HO– | 2.85 |
| C2H5– | 10.30 | F3C– | 5.02 | CH3CONH– | 14.93 |
| (CH3)2CH– | 14.96 | O2N– | 7.63 | CH3CO– | 11.18 |

#####  Other parameters

These can be broadly divided into those that apply to sections of the molecule and those that involve the whole molecule. The former include parameters such as van der Waals’ radii, Charton’s steric constants and the Verloop steric parameters. The latter range from relative molecular mass (RMM) and molar volumes to surface area. They have all been used to correlate biological activity to structure with varying degrees of success.

####  Hansch analysis

Hansch analysis attempts to mathematically relate drug activity to measurable chemical properties. It is based on Hansch’s proposal that drug action could be divided into two stages:

1. the transport of the drug to its site of action;
2. the binding of the drug to the target site.

Each of these stages is dependent on the chemical and physical properties of the drug and its target site. In Hansch analysis these properties are described by the parameters discussed in sections 4.4.1, 4.4.2 and 4.4.3 as well as other parameters. Hansch postulated that the biological activity of a drug could be related to these parameters by simple mathematical relationships based on the general format:

where *C* is the minimum concentration required to cause a specific biological response and *k*1*, k*2*, k*3 and *k*4 are numerical constants obtained by feeding the values of the parameters selected by the investigating team into a suitable computer statistical package. These parameter values are obtained either from the literature (e.g. p, s and *E*s) or determined by experiment (e.g. *C, P* etc.). In investigations where more than one substituent is changed, the value of a specific parameter may be expressed in the Hansch equation as either the sum of the values of that parameter for the individual substituents or independent individual parameters. For example, in the hypothetical case of a benzene ring with two substituents X and Y the Hammett constants could be expressed in the Hansch equation as either *k*1 (sX þ sY) or *k*1sX þ *k*2sY. The equations obtained from the selected data are commonly referred to as Hansch equations. Their precise nature varies (Table 4.8), but for an investigation using *P,* s and *E*s Hansch equations often takes the general form:

P

log 1=*C* ¼ *k*1*P* *k*2*P*2 þ *k*3s þ *k*4*E*S þ *k*5 (4:11)

Parameters other than those shown in equation (4.11) may be used to derive Hansch equations. A comprehensive list may be found in a review by Tute in *Advances in Drug Research* 1971, 6, 1.

The accuracy of a Hansch equation will depend on:

1. the number of analogues (*n*) used: the greater the number the higher the probability of obtaining an accurate Hansch equation;
2. the accuracy of the biological data used in the derivation of the equation. The degree of variation normally found in biological measurements means that a statistically viable number of measurements should be taken for each ana- logue and an average value used in the derivation of the Hansch equation;
3. the choice of parameter (see ‘Craig plots’ below).

The accuracy of a Hansch equation may be assessed from the values of the standard deviation (*s*) and the regression constant (*r*) given by the statistical package used to obtain the equation. The smaller the value of *s* the better the data fits the equation. Values of *r* that are significantly lower than 0.9 indicate that either unsuitable parameter(s) were used to derive the equation or there is no relationship between the compounds used and their activity. This suggests that the mechanisms by which these compounds act are unrelated because the mechanisms are very different from each other.

Hansch equations may be used to predict the activity of an as yet unsynthe- sized analogue. This enables the medicinal chemist to make an informed choice as to which analogues are worth synthesizing. However, these predictions should only be regarded as valid if they are made within the range of parameter values used to establish the Hansch equation. Furthermore, when the predicted activity is widely different from the observed value, it indicates that the activity is affected by factors, such as the ease of metabolism, that were not included in the derivation of the Hansch equation.

Hansch analysis may also be used to give an indication of the importance of the influence of a parameter on the mechanism by which a drug acts. Consider, for example, a series of analogues whose activity is related to the parameters pand s by the hypothetical Hansch equation:

log 1=*C* ¼ 1:78p 0:12s þ 1:674 (4:12)

The small value of the coefficient for s relative to that of p in equation (4.12) shows that the electronic effects do not play an important part in the action of the drug.

##### Craig plots

Craig plots are two dimensional plots of one parameter against another (Figure 4.6). The plot is divided into four sections corresponding to the positive and negative values of the parameters. They are used, in conjunction with an already established Hansch equation for a series of related aromatic compounds, to select the aromatic substituents that are likely to produce highly active



1.00

.CF3SO2

.NO2

SO2NH2

. .

CH3SO2

CN

0.75

. .

.

SF5 .

CONH2

.

CH3CO

0.50 . COOCH3

CF3

COOH

.

.OCF3

0.25

1.2

0.8 0.4

.

CH CONH

3

.F

2.0

1.6

Cl . Br. SCH3

.

I.

0.25 .

OH

0.50

0.4 . 0.8

CH3

.

1.2 1.6

2.0

C2H5

.nC H

4 9

.

OCH3

. NH

2

.N(CH3)2

0.75



An example of a Craig plot of para Hammett constants s against para p values. [Reprinted with permission of John Wiley and Sons, Inc. from Craig P N (1980). In *Burgers Medicinal Chemistry* (M E Wolff, Ed.) 4th ed., Part 1 p. 343. Wiley, New York. Copyright # [1980 John Wiley and Sons Inc.]

analogues. For example, suppose that a Hansch analysis carried out on a series of aromatic compounds yields the Hansch equation:

log 1=C ¼ 2:67p 2:56s þ 3:92 (4:13)

To obtain a high value for the activity (1/C) it is necessary to pick substituents with a positive p value and a negative s value. In other words, if high activity analogues are required, the substituents should be chosen from the lower right- hand quadrant of the plot. However, it is emphasized that the use of a Craig plot does not guarantee that the resultant analogues will be more active than the lead because the parameters used may not be relevant to the mechanism by which the analogue acts.

###  The Topliss decision tree

The Topliss decision tree is essentially a flow diagram that in a series of steps directs the medicinal chemist to produce a series of analogues, some of which should have a greater activity than the lead used to start the tree. It is emphasized that only some of the compounds will be more active than the lead compound. The method is most useful when it is not possible to make the large number of compounds necessary to produce an accurate Hansch equation. However, its use is limited because it requires the lead compound to have an unfused aromatic ring system and it only produces analogues that are substituents of that aromatic system. In addition, the Topliss method also depends on the user being able to rapidly measure the biological activity of the lead compound and its analogues.

There are two Topliss decision trees (Figure 4.7), one for substituents directly attached to an aromatic ring and the other for changes in the aliphatic side chains of an aromatic ring system. Both are used in a similar manner. In both cases the investigation starts with the conversion of the lead into the first analogue at the top of the tree, either the 4-chloro analogue (Figure 4.7(a) ) or the isopropyl analogue (Figure 4.7(b) ). The activity of this analogue is measured and classified as either less (L), approximately the same (E) or significantly greater (M) than that of the original lead. If the activity is greater than that of the lead the next analogue to be prepared is the next one on the M route. Alternatively, if the activity of the analogue is less than that of the original lead the next step is to produce the analogue indicated by the L route on the tree. Similarly, if the activity is about the same as that of the original lead the E route is followed and the appropriate analogue synthesized. This procedure is repeated, the activity of each new analogue being compared with that of its precursor in order to determine which branch of the tree gives the next ana- logue.