**Lecture IV**

**Drug metabolism. Structure-activity relationships.**

# Most metabolic products are less pharmacologically active

# Important exceptions:

Where the metabolite is more active (Prodrugs, e.g. Erythromycin-succinate (less irritation of GI) --> Erythromycin)

Where the metabolite is toxic (acetaminophen)

Where the metabolite is carcinogenic

Close relationship between the biotransformation of drugs and normal biochemical processes occurring in the body:

* Metabolism of drugs involves many pathways associated with the synthesis of endogenous substrates such as steroid hormones, cholesterol and bile acids
* Many of the enzymes involved in drug metabolism are principally designed for the metabolism of endogenous compounds
* These enzymes metabolize drugs only because the drugs resemble the natural compound

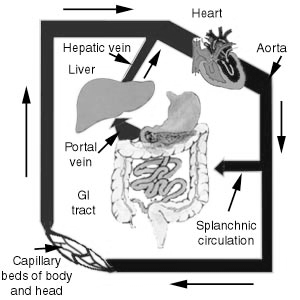
# **Phases of Drug Metabolism**

### Phase I Reactions

* Convert parent compound into a more polar (=hydrophilic) metabolite by adding or unmasking functional groups (-OH, -SH, -NH2, -COOH, etc.)
* Often these metabolites are inactive
* May be sufficiently polar to be excreted readily

Phase II Reactions

* Conjugation with endogenous substrate to further increase aqueous solubility
* Conjugation with glucoronide, sulfate, acetate, amino acid
* Phase I usually precede phase II reactions

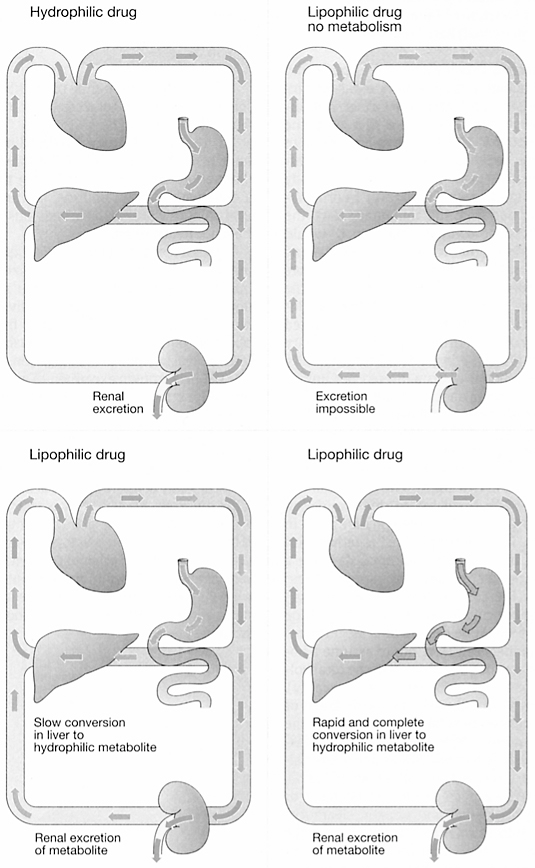
 Liver is principal site of drug metabolism:

* Other sites include the gut, lungs, skin and kidneys
* For orally administered compounds, there is the

“First Pass Effect”

* Intestinal metabolism
* Liver metabolism
* Enterohepatic recycling
* Gut microorganisms – glucuronidases

**Drug Metabolism**



**Drug Metabolism - Phase I**

**Phase I Reactions**

* Oxidation
* Reduction
* Hydrolytic cleavage
* Alkylation (Methylation)
* Dealkylation
* Ring cyclization
* N-carboxylation
* Dimerization
* Transamidation
* Isomerization
* Decarboxylation

# **Drug Metabolism - Oxidation**

### Two types of oxidation reactions:

### Oxygen is incorporated into the drug molecule (e.g. hydroxylation)

### Oxidation causes the loss of part of he drug molecule (e.g. oxidative deimination, dealkylation)

Microsomal Mixed Function Oxidases (MFOs)

“Microsomes” form in vitro after cell homogenization and fractionation of ER

* Rough microsomes are primarily associated with protein synthesis
* Smooth microsomes contain a class of oxidative enzymes called

### “Mixed Function Oxidases” or “Monooxygenases”

### These enzymes require a reducing agent (NADPH) and molecular oxygen (one oxygen atom appearing in the product and the other in the form of water)

# Drug Metabolism - Oxidation

## MFO consists of two enzymes:

## Flavoprotein, NADPH-cytochrome c reductase

## One mole of this enzyme contains one mole each of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD)

## Enzyme is also called NADPH-cytochrome P450 reductase

### Cytochrome P450

### named based on its light absorption at 450 nm when complexed with carbon monoxide

### is a hemoprotein containing an iron atom which can alternate between the ferrous (Fe++) and ferric (Fe+++) states

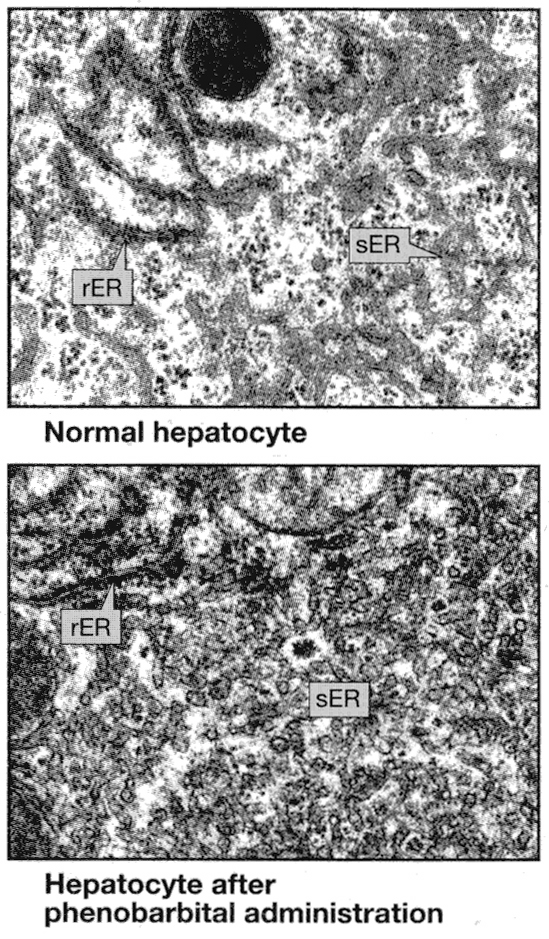
### Electron acceptor

### Serves as terminal oxidase its relative abundance compared to NADPH-cytochrome P450 reductase makes it the rate-limiting step in the oxidation reactions

# **Drug Metabolism - Oxidation**

* + Humans have 18 families of cytochrome P450 genes and 43 subfamilies:
    - CYP1 drug metabolism (3 subfamilies, 3 genes, 1 pseudogene)
    - CYP2 drug and steroid metabolism (13 subfamilies, 16 genes, 16 pseudogenes)
    - **CYP3 drug metabolism** (1 subfamily, 4 genes, 2 pseudogenes)
    - CYP4 arachidonic acid or fatty acid metabolism (5 subfamilies, 11 genes, 10 pseudogenes)
    - CYP5 Thromboxane A2 synthase (1 subfamily, 1 gene)
    - CYP7A bile acid biosynthesis 7-alpha hydroxylase of steroid nucleus (1 subfamily member)
    - CYP7B brain specific form of 7-alpha hydroxylase (1 subfamily member)
    - CYP8A prostacyclin synthase (1 subfamily member)
    - CYP8B bile acid biosynthesis (1 subfamily member)
    - CYP11 steroid biosynthesis (2 subfamilies, 3 genes)
    - CYP17 steroid biosynthesis (1 subfamily, 1 gene) 17-alpha hydroxylase
    - CYP19 steroid biosynthesis (1 subfamily, 1 gene) aromatase forms estrogen
    - CYP20 Unknown function (1 subfamily, 1 gene)
    - CYP21 steroid biosynthesis (1 subfamily, 1 gene, 1 pseudogene)
    - CYP24 vitamin D degradation (1 subfamily, 1 gene)
    - CYP26A retinoic acid hydroxylase important in development (1 subfamily member)
    - CYP26B probable retinoic acid hydroxylase (1 subfamily member)
    - CYP26C probabvle retinoic acid hydroxylase (1 subfamily member)
    - CYP27A bile acid biosynthesis (1 subfamily member)
    - CYP27B Vitamin D3 1-alpha hydroxylase activates vitamin D3 (1 subfamily member)
    - CYP27C Unknown function (1 subfamily member)
    - CYP39 7 alpha hydroxylation of 24 hydroxy cholesterol (1 subfamily member)
    - CYP46 cholesterol 24-hydroxylase (1 subfamily member)

CYP51 cholesterol biosynthesis (1 subfamily, 1 gene, 3 pseudogenes) lanosterol 14-alpha demethylase

**Drug Metabolism - Oxidation**

Induction of P450 enzymes:

PPAR (peroxisome proliferator activated receptor) ligands(e.g.clofibrate). CYP1 family are induced by aromatic hydrocarbons (cigarette smoke; charred food). CYP2E enzymes induced by ethanol. CYP2B enzymes induced 40-50 fold by barbiturates.

Polymorphisms cause differences in drug metabolism:

CYP2C19 has a polymorphism that changes the enzyme's ability to metabolize mephenytoin (a marker drug). In Caucasians, the polymorphism for the poor metabolizer phenotype is only seen in 3% of the population. However, it is seen in 20% of the asian population. It is İmportant to be aware of a person's race when drugs are given that are metabolized differently by different populations.

P450s and drug interactions:

* Barbiturates induce CYP2B => increased metabolism of other drugs
* Antifungals (e.g. ketoconazole) inhibit fungal CYP51 and unintentionally also human CYP3A4 => reduced metabolism of other drugs
* Grapefruit juice contains a CYP3A4 inhibitor =>12 fold increase in some drug concentrations
* CYP3A4 Substrates: • Acetominophen (Tylenol) • Codeine (narcotic) • Cyclosporin A (immunosuppressant),

Diazepam (Valium) • Erythromycin (Antibiotic) • Lidocaine (local anaesthetic), • Lovastatin (HMGCoA reductase inhibitor), • Taxol (cancer drug), • Warfarin (anticoagulant).

# **Drug Metabolism - Oxidation**

### Drug oxidation requires:

* + Cytochrome P450
  + Cytochrome P450 reductase
  + NADPH
  + Molecular oxygen

### The cycle involves four steps:

### Oxidized (Fe3+) cytochrome P-450 combines with a drug substrate to

### form a binary complex.

### NADPH donates an electron to the cytochrome P-450 reductase, which

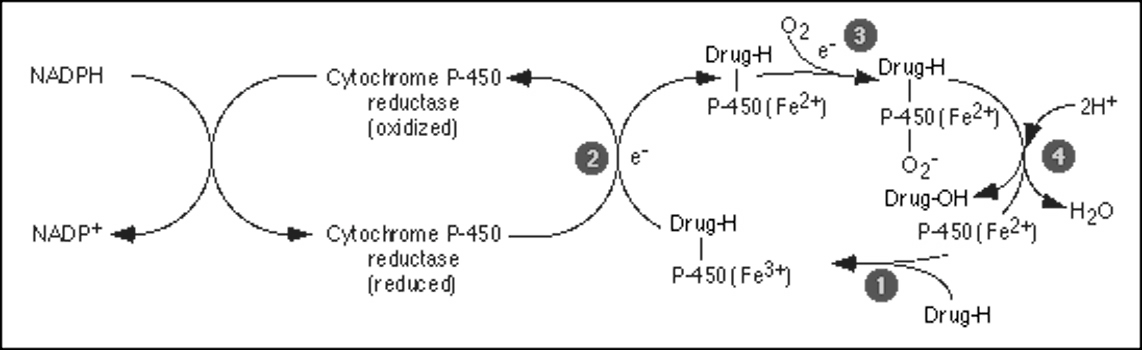
### in turn reduces the oxidized cytochrome P-450-drug complex.

1. A second electron is introduced from NADPH *via* the same cytochrome P-450

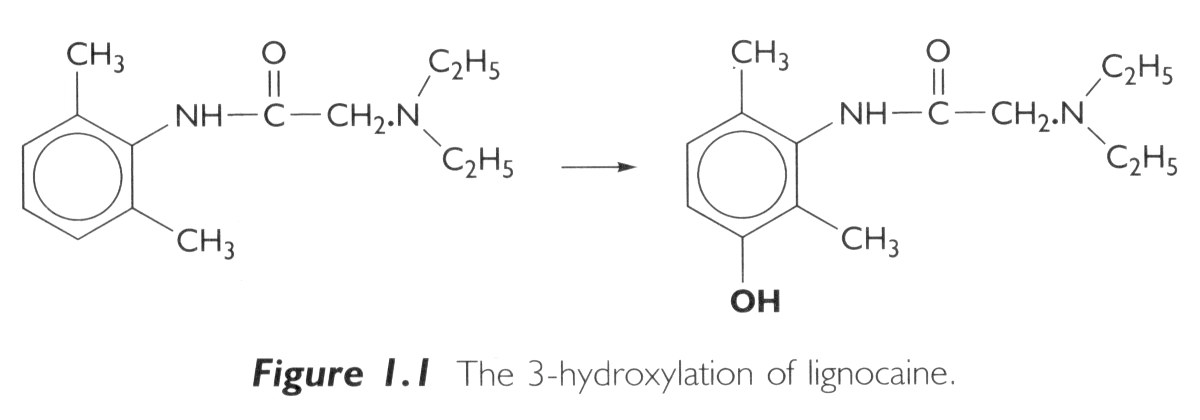
reductase, which serves to reduce molecular oxygen and form an "activated oxygen"-cytochrome P-450-substrate complex.

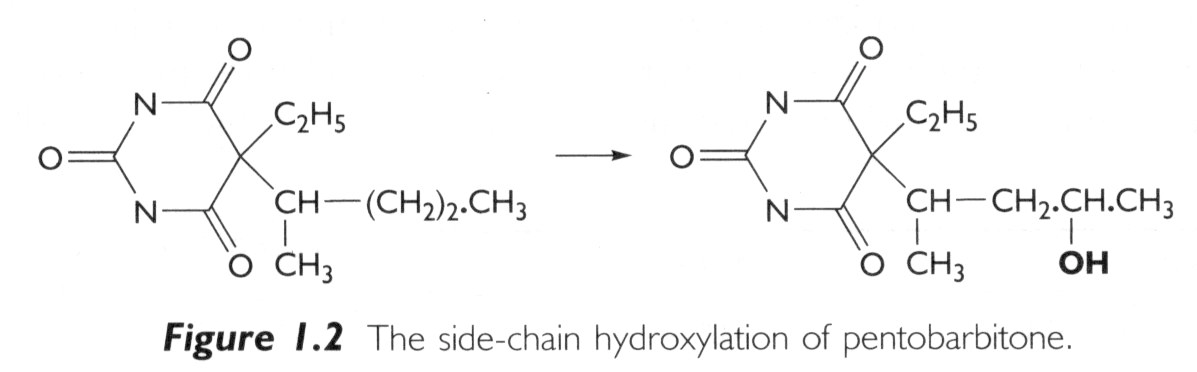
1. This complex in turn transfers "activated" oxygen to the drug substrate to

form the oxidized product. The potent oxidizing properties of this activated oxygen permit oxidation of a large number of substrates.

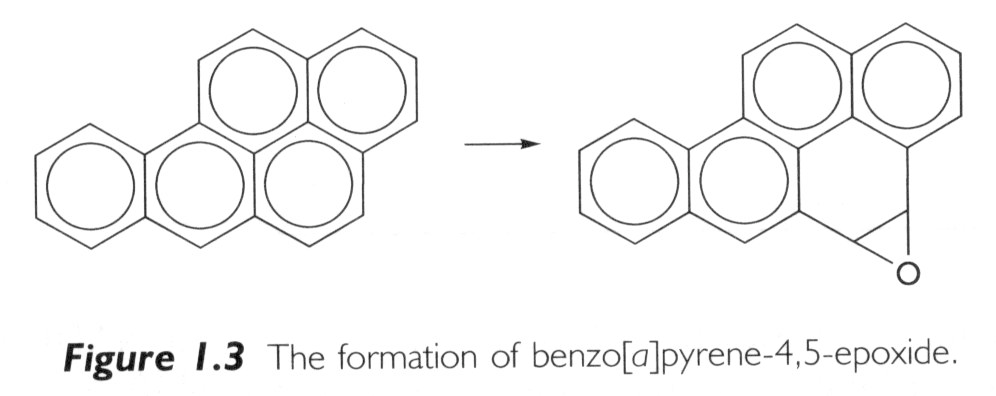
**Drug Metabolism - Oxidation**

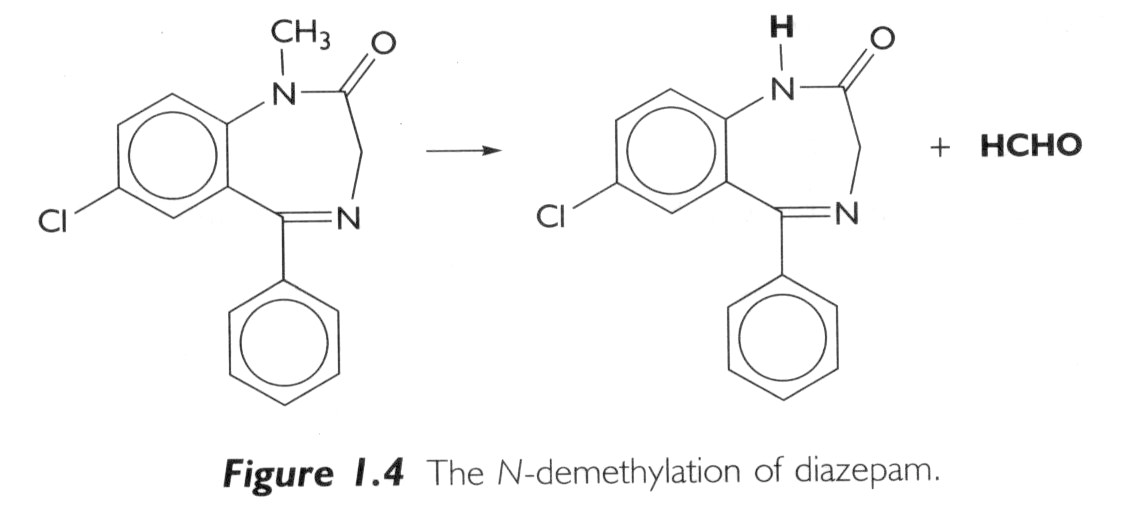
Aromatic hydroxylation:



Aliphatic hydroxylation:

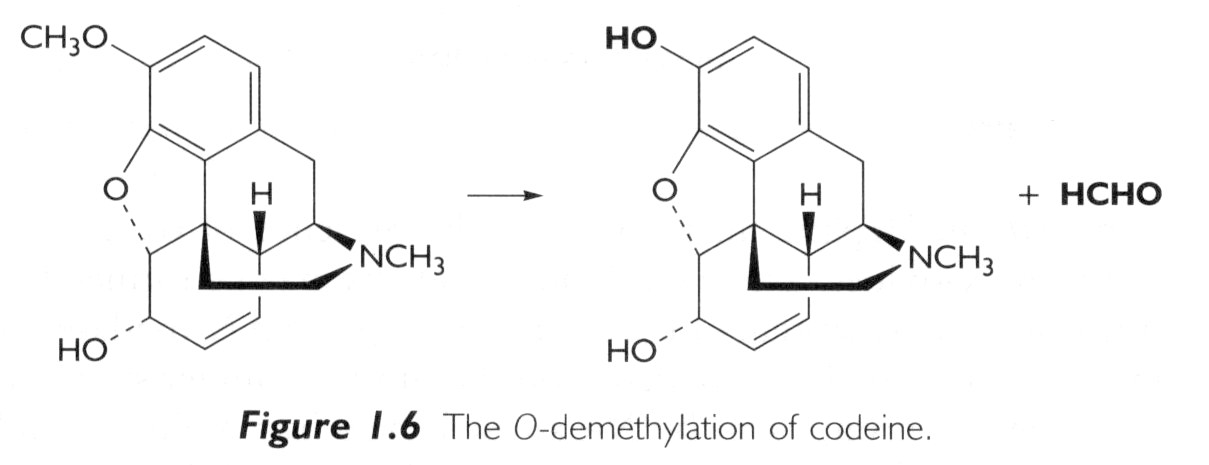
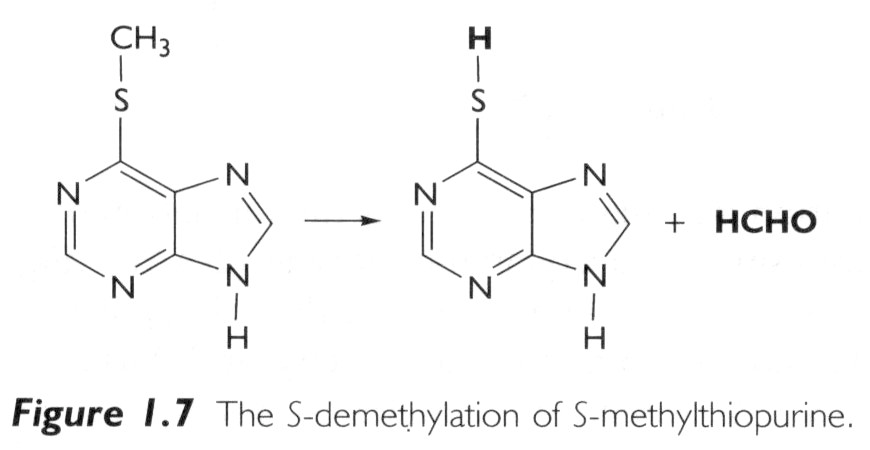
**Drug Metabolism - Oxidation**

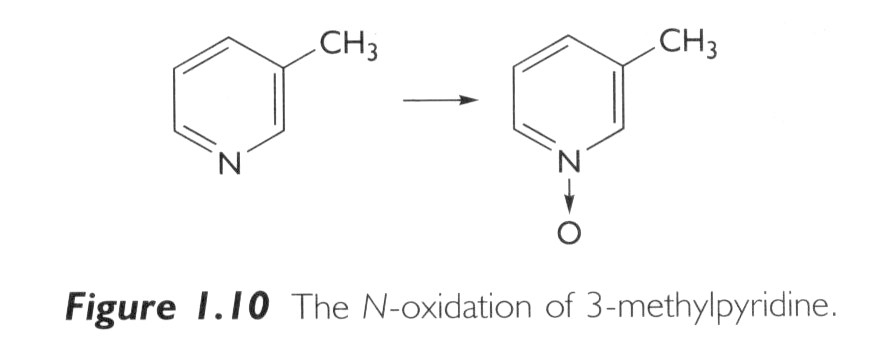
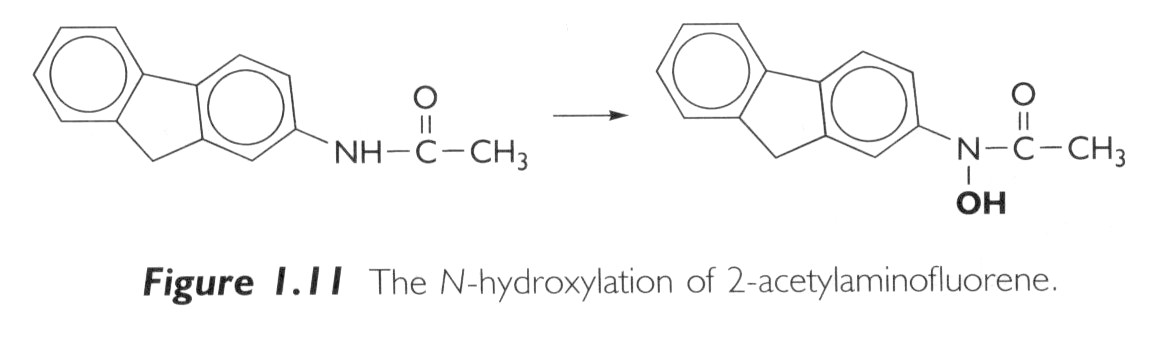
Epoxidation:

Dealkylation:

**Drug Metabolism - Oxidation**

O-demethylation: S-demethylation:



N-oxidation: N-hydroxylation:

**Drug Metabolism – Oxidation**

Oxidation reactions NOT catalyzed by Cytochrome P450:

Flavin containing monoxygenase system

* Present mainly in liver but some is expressed in gut and lung
* Located in smooth endoplasmic reticulum
* Oxidizes compounds containing sulfur and nitrogen
* Uses NADH and NADPH as cofactors
* Alcohol dehydrogenase (cytosol)
* Aldehyde oxidation (cytosol)
* Xanthine oxidase
* Amine oxidases
* Monoamine oxidase (nerve terminals, mitochondria)
* Diamine oxidase found in liver microsomes

Primarily endogenous metabolism

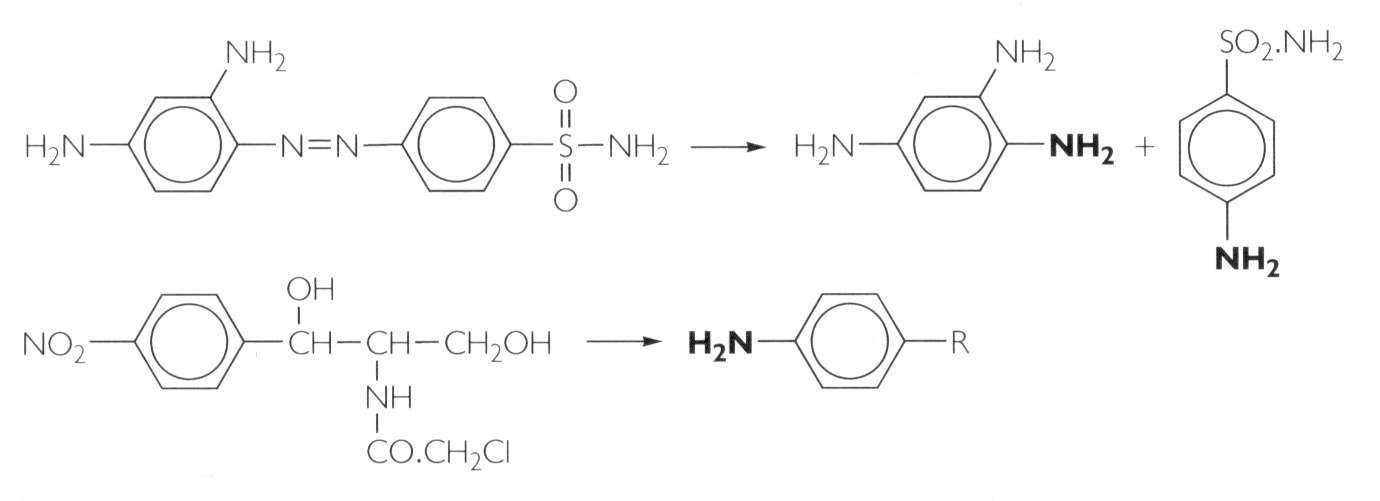
Drug Metabolism – Oxidation

Monoamine Oxidases (MAO):

* Catalyze oxidative deamination of endogenous catecholamines (epinephrine)
* Located in nerve terminals and peripheral tissues
* Substrates for catecholamine metabolism found in foods (tyramine) can cause a drug/food interaction
* Inhibited by class of antidepressants called MAO inhibitors

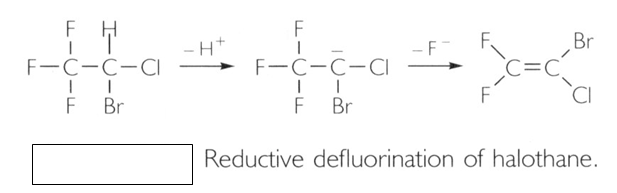
(Inhibition of MAO isoforms in the CNS also effects levels of serotonin - Tranylcypromine)

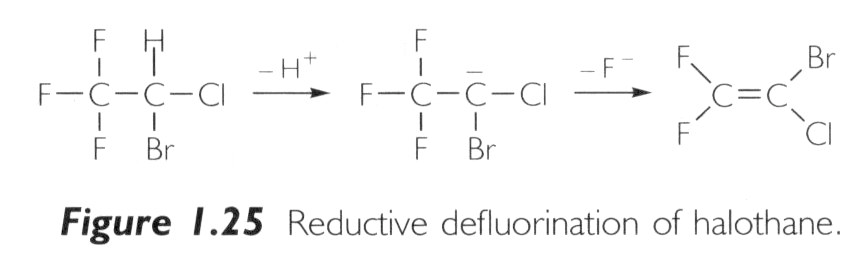
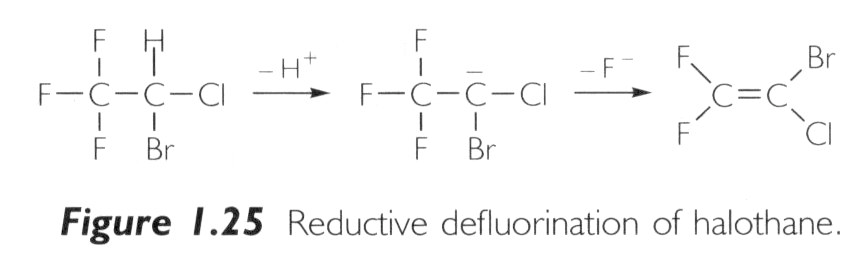
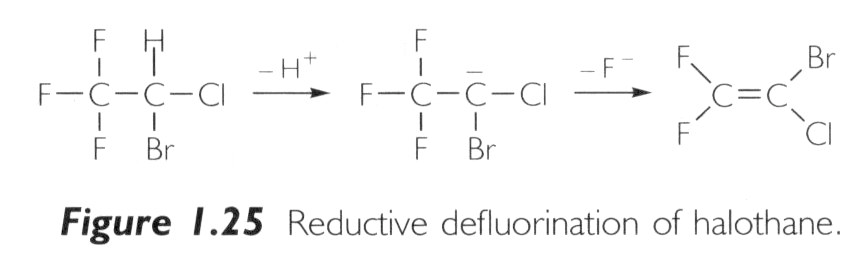
These drugs can cause severe or fatal drug/drug interactions with drugs that increase release of catecholamines or inhibit their reuptake in nerve terminals (Meperidine, pentazocine, dextromethorphan, SSRI antidepressants).

**Drug Metabolism - Reduction**

Azo-reduction:

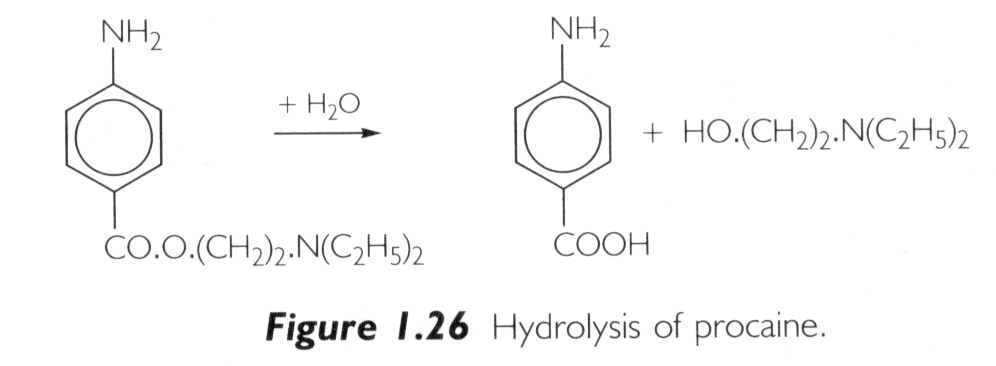
Nitro-reduction:

Dehalogenation: 

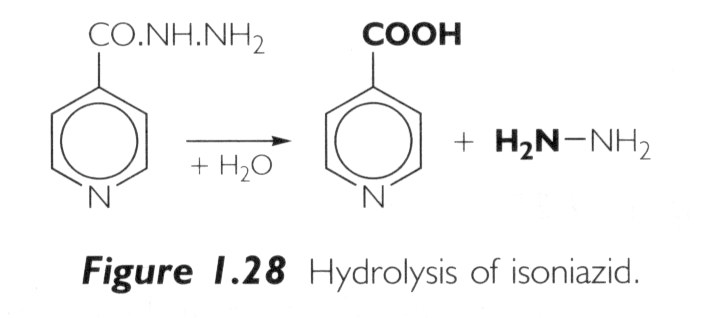


**Drug Metabolism - Reduction**

**Hydrolysis reactions**

****

**Ester hydrolysis**:



**Amide hydrolysis:**

**Drug Metabolism - Phase I**

* Almost any drug can undergo modifications by drug-metabolizing enzyme systems
* Drugs can be subject to several Phase I pathways
* These reactions create functional groups that place the drugs in a correct chemical state to be acted upon by Phase II conjugative mechanisms
* Main function of phase I reactions is to prepare chemicals for phase II metabolism and subsequent excretion
* Phase II is the true “detoxification” step in the metabolism process.

**Drug Metabolism - Phase II**

**Conjugation reactions**

– Glucuronidation by UDP-Glucuronosyltransferase:

(on -OH, -COOH, -NH2, -SH groups)

– Sulfation by Sulfotransferase:

(on -NH2, -SO2NH2, -OH groups)

– Acetylation by acetyltransferase:

(on -NH2, -SO2NH2, -OH groups)

– Amino acid conjugation

(on -COOH groups)

– Glutathione conjugation by Glutathione-S-transferase:

(to epoxides or organic halides)

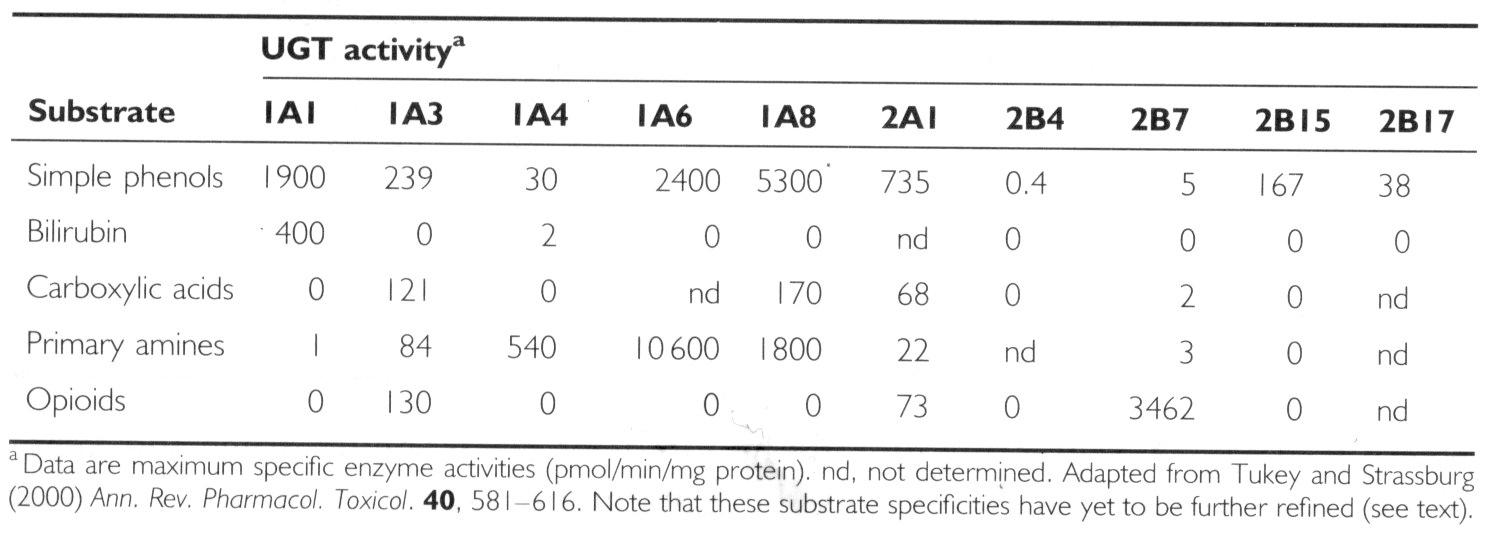
– Fatty acid conjugation

(on -OH groups)

– Condensation reactions

**Drug Metabolism - Glucuronidation**

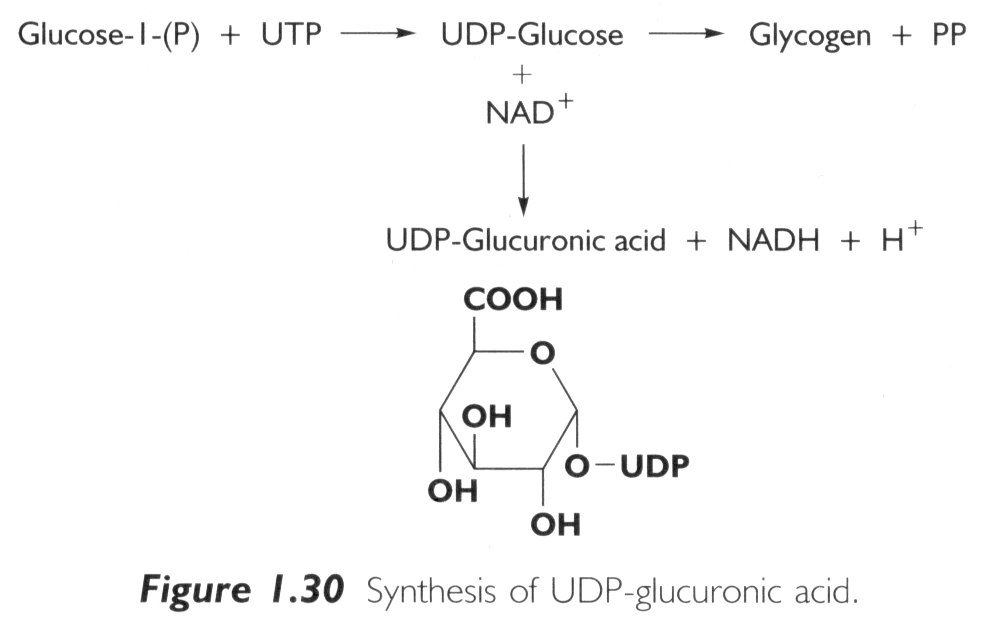
* Glucuronidation ( = conjugation to -d-glucuronic acid)
* Quantitatively the most important phase II pathway for drugs and endogenous compounds
* Products are often excreted in the bile.
* Enterohepatic recycling may occur due to gut glucuronidases
* Requires enzyme UDP-glucuronosyltransferase (UGT):
* Genetic family of enzymes
* Metabolizes a broad range of structurally diverse endogenous and exogenous compounds

Structurally related family with approximately 16 isoforms in man

**Drug Metabolism - Glucuronidation**

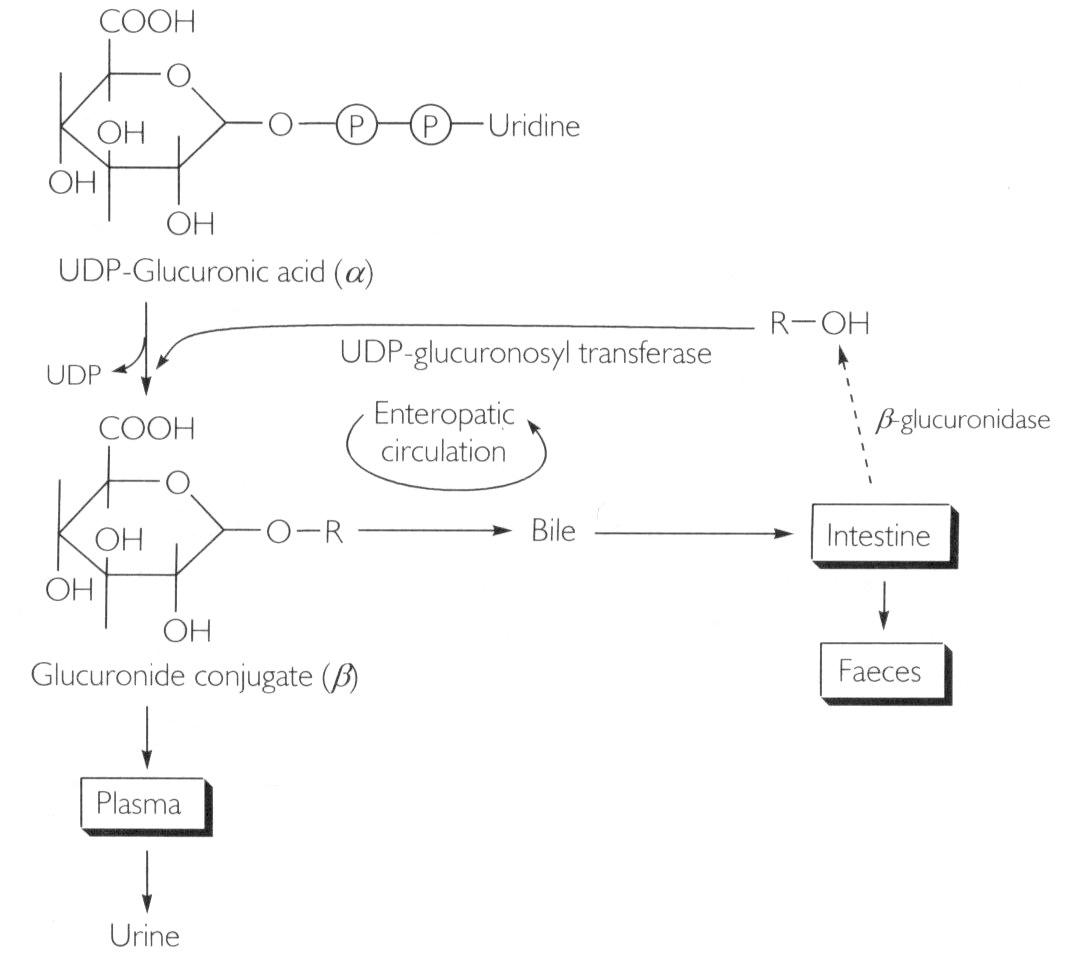
* Glucuronidation – requires creation of high energy intermediate:

UDP-Glucuronic Acid:

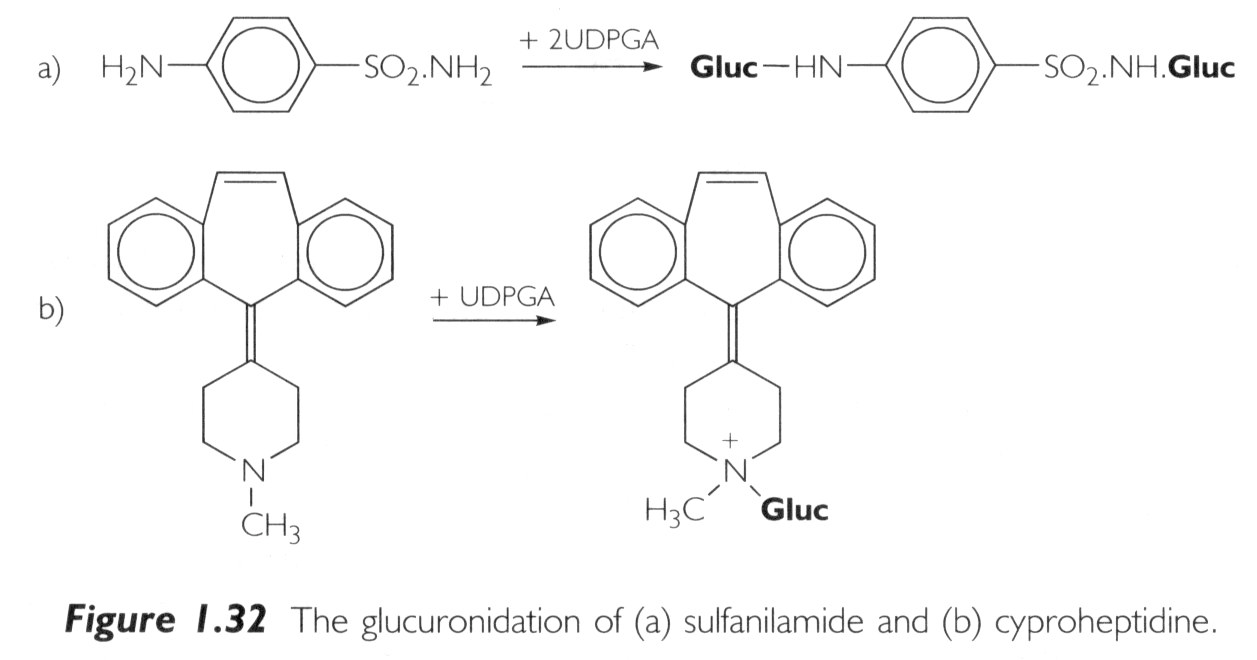


**Drug Metabolism – Glucuronidation**

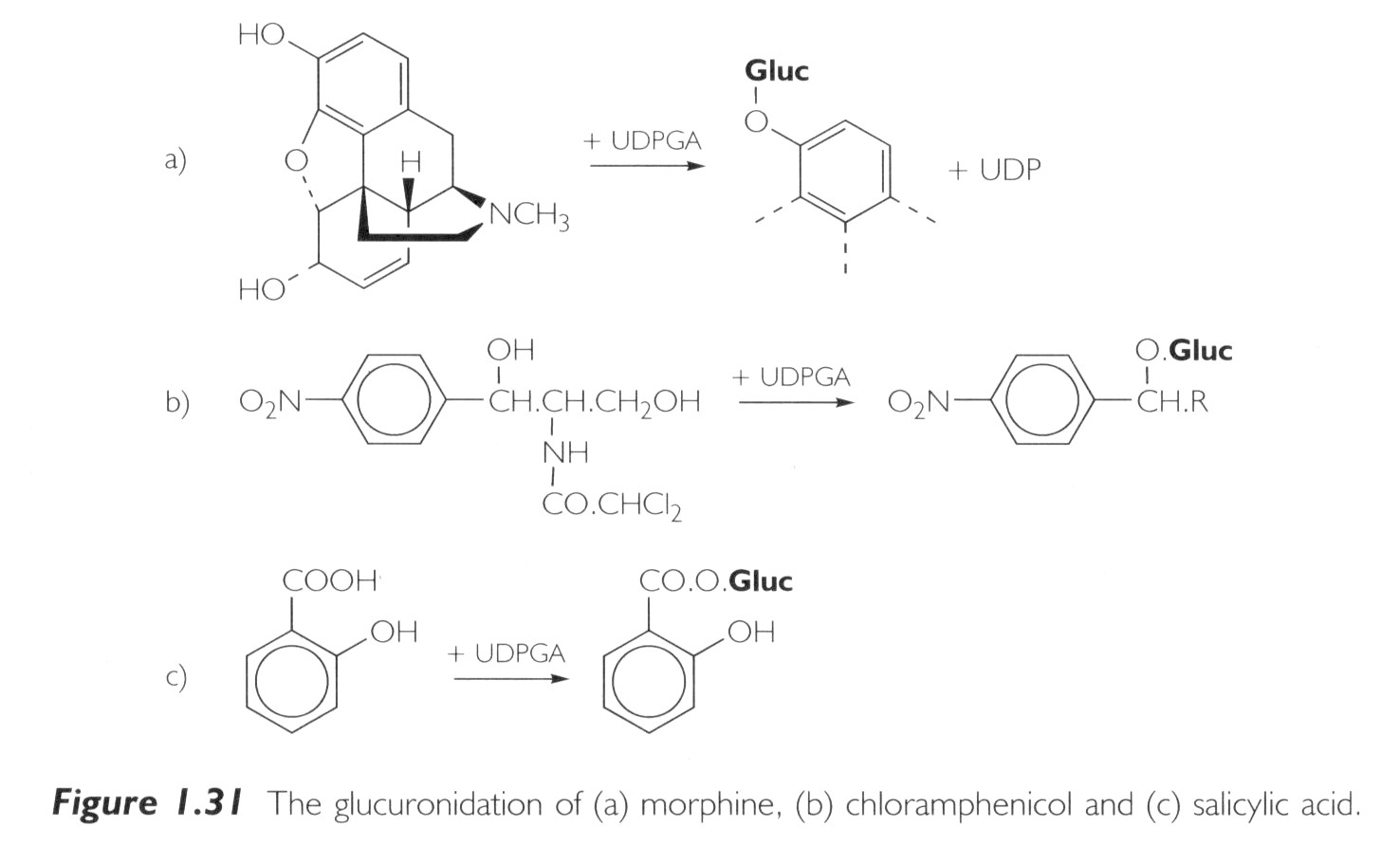
Glucuronidation Pathway and Enterohepatic Recirculation



**Drug Metabolism - Glucuronidation**

* N-glucuronidation:
* Occurs with amines (mainly aromatic )
* Occurs with amides and sulfonamides

**Drug Metabolism - Glucuronidation**

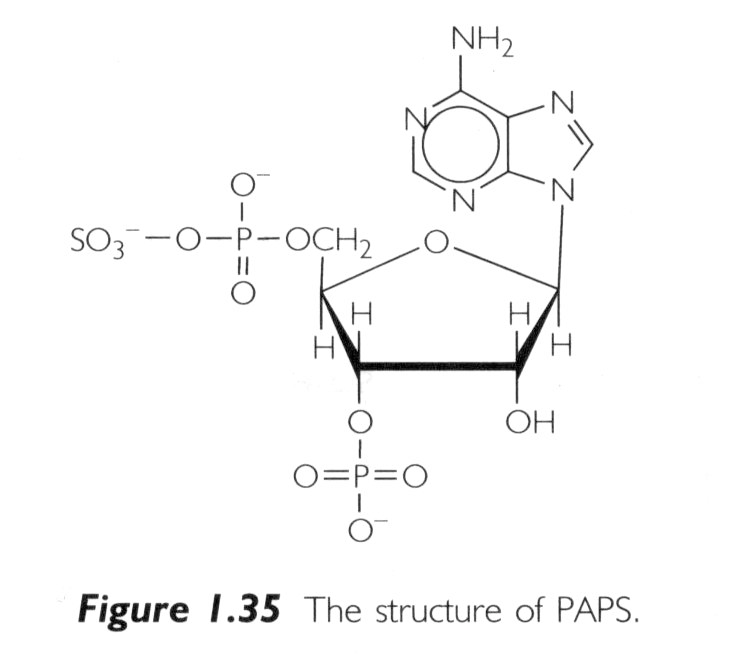
* O-glucuronidation:
  + Occurs by ester linkages with carboxylic acids
  + Occurs by ether linkages with phenols and alcohols

**Drug Metabolism - Sulfation**

Sulfation:

* Major pathway for phenols but also occurs for alcohols, amines and thiols
* Energy rich donor required:

PAPS (3’-Phosphoadenosine-5’-phosphosulfate)



Sulfation and glucuronidation are competing pathways:

* + Sulfation predominates at low substrate concentrations
  + Glucuronidation predominates at higher concentrations

There is relatively less PAPS in cell cytosol compared to UDPGA

* Sulfotransferases (=SULTs) catalyze transfer of sulfate to substrates:
  + Phenol, alcohol and arylamine sulfotransferases are fairly non-specific

Steroid sulfotransferases are very specific

**Drug Metabolism – Acylation**

**Acetylation:**

* Common reaction for aromatic amines and sulfonamides
* Requires co-factor acetyl-CoA
* Responsible enzyme is N-acetyltransferase
* Takes place mainly in the liver
* Important in sulfonamide metabolism because acetyl-sulfonamides are less soluble than the parent compound and may cause renal toxicity due to precipitation in the kidney

**Fatty Acid Conjugation:**

* Stearic and palmitic acids are conjugated to drug by esterification reaction
* Occurs in liver microsomal fraction

**Drug Metabolism - Other conjugations**

**Amino Acid Conjugation:**

• ATP-dependent acid:CoA ligase forms active CoA-amino acid conjugates which then react with drugs by N-Acetylation:

– Usual amino acids involved are:

• Glycine. Glutamine, Ornithine, Arginine

**Glutathione Conjugation:**

• Tripeptide Gly-Cys-Glu; conjugated by glutathione-S-transferase (GST)

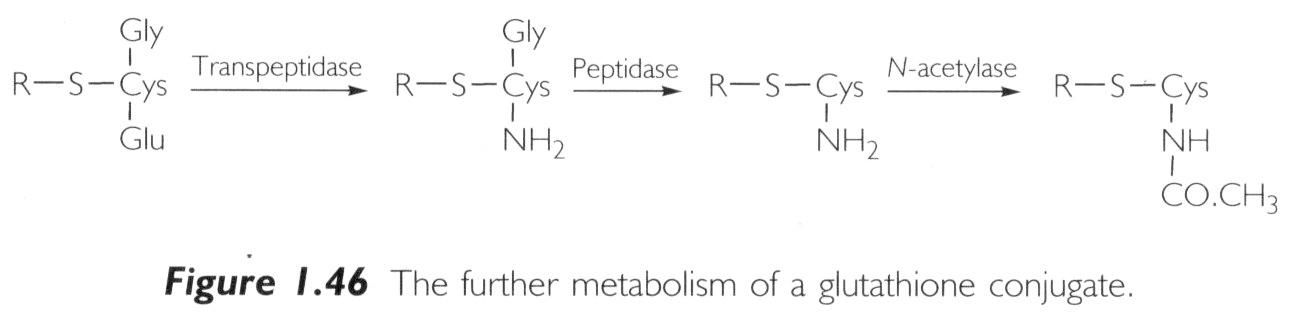
• Glutathione is a protective factor for removal of potentially toxic compounds

• Conjugated compounds can subsequently be attacked by

-glutamyltranspeptidase and a peptidase to yield the cysteine conjugate => product can be further acetylated to N-acetylcysteine conjugate

**Drug Metabolism - Phase I & II**

Phase I and II - Summary:



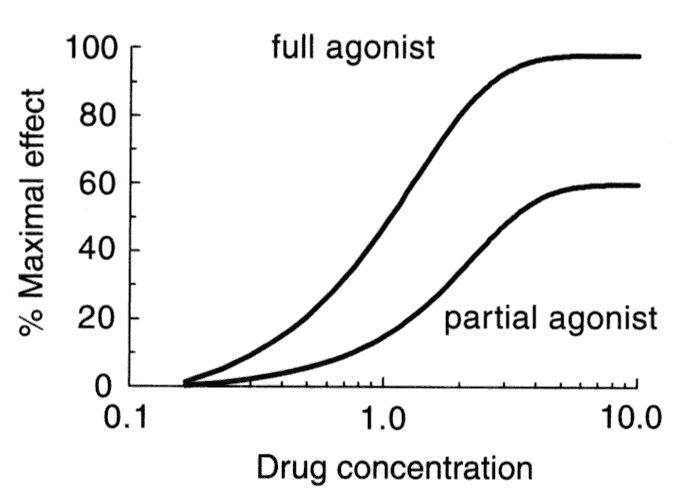
* Products are generally more water soluble
* These reactions products are ready for (renal) excretion
* There are many complementary, sequential and competing pathways

Phase I and Phase II metabolism are a coupled interactive system interfacing with endogenous metabolic pathways

**Drug Action: Receptor Theory**

Many drugs act by binding to receptors (see Lecture 4) where they either provoke or inhibit a biological response.

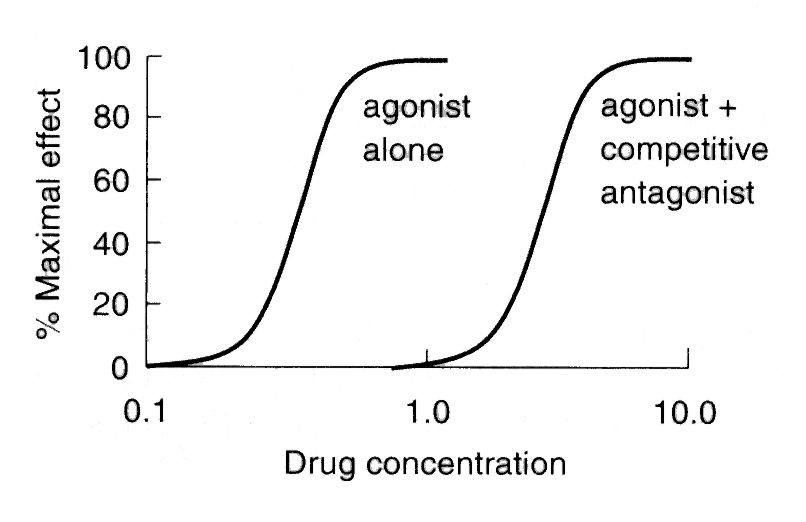
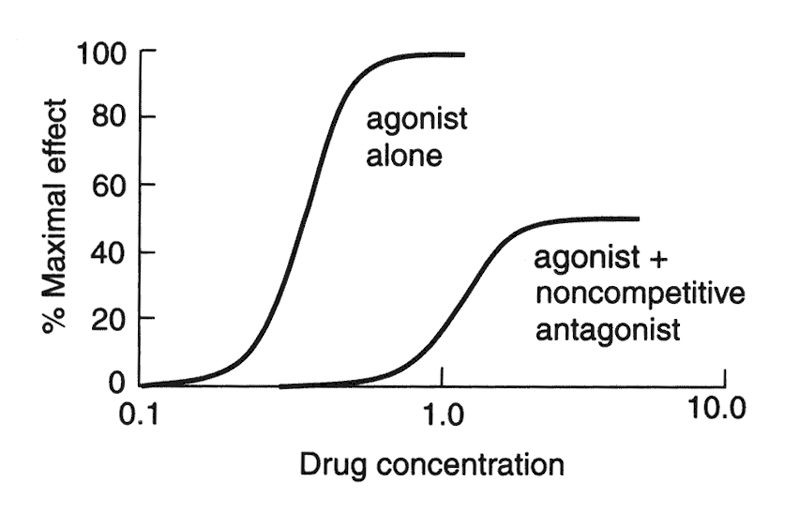
Agonists:

* Can be drugs or endogenous ligands for the receptor
* Increasing concentrations of the agonist will produce an increase in the biological response:
* Full Agonist: Evokes 100% of the maximum possible effect
* Partial Agonist: Produces the same type of biological response, but cannot achieve 100% even at very high doses

**Drug Action: Receptor Theory**

**Antagonists:**

* Block or reverse the effects of agonists. They have no effects on their own
* Competitive Antagonists: Compete with agonist for receptor binding => Agonist appears less potent, but can still achieve 100% effect (but at higher concentrations)
* Non-competitive Antagonists: Bind to receptor at different site and either prevent agonist binding or the agonist effect => maximal achievable response reduced
* Inverse Agonists: Not the same as antagonists! Inverse agonists trigger a negative response (= reduce baseline) (e.g. diazepam = full agonist = anticonvulsant BUT inverse agonists of benzodiazepin receptor are convulsants)



## **The SAR and QSAR approaches to drug design**

### **Structure–activity relationships (SARS)**

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)**

Figure 4.1 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 (Table 4.1). 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 (Figure 4.2).

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 (see Chapter 9). 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 reduced hepatotoxicity. It is believed that this reduction is due to the methyl groups preventing metabolic hydroxylation of these ortho positions.

Steric hindrance between the hydrogen atom and the lone pairs.

O



H H C

H**..**

O**..**

N N

Diphenhydramine o-Methyl analogue

Figure 4.2 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.

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 in their chemical and/or physical properties (Table 4.3). As a result, they may exhibit similar pharmacokinetic and pharmacodynamic properties. In other words, the replacement of a substituent by its isostere is more likely to result in the formation of an analogue with the same type of activity as the lead than the totally random selection of an alternative substituent. However, luck still plays a part, and an isosteric analogue may have a totally different type of activity from its lead

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

Figure 4.3 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

A large number of drugs have been discovered by isosteric interchanges (Figure 4.4).

OH H SH H S

N

N

N

N

N N

N N N N R R

Hypoxanthine 6-Mercaptopurine

(Antitumour agent)

Phenothiazine drugs (Neuroleptics)

Dibenzazepine drugs (Neuroleptics)

Figure 4.4 Examples of drugs discovered by isosteric replacement

### 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:

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

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 (Table 4.4). 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:

sX ¼ log (*K*BX=*K*B) (4:5)

that is

and so, as p*K*a ¼ log *K*a,

sX ¼ log *K*BX log *K*B (4:6)

sX ¼ p*K*B p*K*BX (4:7)

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.

Table 4.5 Examples of the different electronic substitution constants used in QSAR studies. Inductive substituent constants (s1) are the contribution the inductive effect makes to Hammett constants and can be used for aliphatic compounds. Taft substitution constants (s\*) refer to aliphatic substituents but use propanoic acid (the 2-methyl derivative of ethanoic acid) as the reference point. The Swain–Lupton constants represent the contributions due to the inductive (*F* ) and mesomeric or resonance (*R*) components of Hammett constants. Adapted from *An Introduction to the Principles of Drug Design and Action* by Smith and Williams 3rd Ed. (1998) Ed. H.J.Smith. Reproduced by permission of Harwood Academic Publishers.

Substituent Hammett constants Inductive constants Taft constants Swain–Lupton constants

sm sp s1 s\* *F R*

–H 0.00 0.00 0.00 0.49 0.00 0.00

–CH3 0.07 0.17 0.05 0.00 0.04 0.13

–C2H5 0.07 0.15 0.05 0.10 0.05 0.10

–Ph 0.06 0.01 0.10 0.60 0.08 0.08

–OH 0.12 0.37 0.25 — 0.29 0.64

–Cl 0.37 0.23 0.47 — 0.41 0.15

–NO2 0.71 0.78 — — 0.67 0.16

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:

*k*(XCH2COOCH3)

¼ ¼*k*

*E*s log log *k*

(CH COOCH )

(XCH2COOCH3)

— log *k*(CH3COOCH3)

(4:8)

3 3

where *k* is the rate constant of the appropriate hydrolysis and the value of *E*s ¼ 0 when X ¼ CH3. It is assumed that the values for *E*s (Table 4.6.) obtained for a group using the hydrolysis data are applicable to other structures containing that group. The methyl based *E*s values can be converted to H based values by adding 1.24 to the corresponding methyl based values.

Table 4.6 Examples of the Taft steric parameter *E*s

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Group | *E*s | Group | *E*s | Group | *E*s |
| H– | 1.24 | F– | 0.78 | CH3O– | 0.69 |
| CH3– | 0.00 | Cl– | 0.27 | CH3S– | 0.19 |
| C2H5– (CH3)2 CH– | 0.07  0.47 | F3C– Cl3C– | 1.16  2.06 | PhCH2– PhOCH– | 0.38  0.33 |

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:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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:

log 1=*C* ¼ *k*1(partition parameter) þ *k*2(electronic parameter) þ *k*3(steric parameter) þ *k*4 (4:10)

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:

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).

Table 4.8 Examples of simple Hansch equations

Compound Activity Hansch equation

X

Br Antiadrenergic log 1=*C* ¼ 1:22p 1:59s þ 7:89

Y CHCH2N(CH3)2 HCl

(CH2) n CH3

(*n* ¼ 22; *s* ¼ 0:238; *r* ¼ 0:918)

OCHCO S

X C N

O

CH3 CH3

Antibiotic (*in vivo*)

log 1=*C* ¼ 0:445p þ 5:673

(*n* ¼ 20; *r* ¼ 0:909)

COOH

X

 OCH2CH2NH Y

X OH

B

OH

MAO

inhibitor (humans)

Concentration (*C*b) in the brain after

15 minutes

log 1=*C* ¼ 0:398p þ 1:089s þ 1:03*E*s þ 4:541 (*n* ¼ 9; *r* ¼ 0:955)Þ

log *C*b ¼ 0:765p 0:540p2 þ 1:505

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 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

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



Figure 4.6 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.]

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. Suppose, for example, that a compound A (Figure 4.8) is active against

*S. aureus* and activity of this compound and its analogues can be readily assessed by a biological method. The first step in the Topliss approach is to synthesize the 4-chloro derivative (B) of A. Suppose the activity of B is greater than that of A, then following the M branch the Topliss tree (Figure 4.8) indicates that the next analogue to produce is the 3,4-dichloro derivative (C) of A. Once again suppose that the biological assay of C was less than that of B. In this case, the Topliss tree shows that the next most promising analogue is the 4-trifluromethyl derivative of (D) of A. At this point one would also synthesize and biologically test the 2,4-dichloro (E) and the 4-nitro analogues (F) of A. It is emphasized that the decision tree is not a synthetic pathway for the production of each of the analogues. It simply suggests which of the substituents would be likely to yield a more potent analogue. The synthetic route for producing each of the suggested analogues would vary for each analogue and would use the most appropriate starting materials.

The Topliss decision tree does not give all the possible analogues but it is likely that a number of the most active analogues will be found by this method.