The Organic Chemistry of Drug Design and Drug Action

Richard B. Silverman
Department of Chemistry
Northwestern University
Evanston, Illinois

ACADEMIC PRESS
A Division of Harcourt Brace & Company
San Diego New York Boston London Sydney Tokyo Toronto

Find Us on the Web! http: //www.apnet.com

This book is printed on acid-free paper. ♾

Academic Press
A Division of Harcourt Brace & Company
525 B Street, Suite 1900, San Diego, California 92101-4495
United Kingdom Edition published by
Academic Press Limited
24–28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Silverman, Richard B.
The organic chemistry of drug design and drug action / Richard B. Silverman.
p. cm.
Includes index.
ISBN 0-12-643730-0 (hardcover)
1. Pharmaceutical chemistry. 2. Bioorganic chemistry. 3. Molecular pharmacology. 4. Drugs--Design. I. Title.
[DNLM: 1. Chemistry, Organic. 2. Chemistry, Pharmaceutical. 3. Drug Design. 4. Pharmacokinetics. QV 744 S587o]
RS403.S55 1992
615'.19--dc20
DNLM/DLC
for Library of Congress

91-47041
CIP

Printed in the United States of America
98 99 MM 8 7 6

CHAPTER 2

Drug Discovery, Design, and Development

I. Drug Discovery 4
  A. Drug Discovery without a Lead 5
    1. Penicillins, 5 • 2. Librium, 6
  B. Lead Discovery 7
    1. Random Screening, 8 • 2. Nonrandom Screening, 9 • 3. Drug Metabolism Studies, 9 • 4. Clinical Observations, 9 • 5. Rational Approaches to Lead Discovery, 10

II. Drug Development: Lead Modification 11
  A. Identification of the Active Part: The Pharmacophore 11
  B. Functional Group Modification 13
  C. Structure–Activity Relationships 14
  D. Structure Modifications to Increase Potency and Therapeutic Index 15
    1. Homologation, 16 • 2. Chain Branching, 18 • 3. Ring–Chain Transformations, 18 • 4. Bioisosterism, 19
  E. Quantitative Structure–Activity Relationships 23
    1. Historical, 23 • 2. Physicochemical Parameters, 24 • 3. Methods Used to Correlate Physicochemical Parameters with Biological Activity, 35 • 4. Computer-Based Methods of QSAR Related to Receptor Binding, 43
  F. Molecular Graphics-Based Drug Design 44
  G. Epilogue 47

References 47
General References 50

I. Drug Discovery

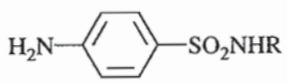
In general, clinically used drugs are not discovered. What is more likely discovered is known as a *lead* compound. The lead is a prototype compound that has the desired biological or pharmacological activity, but may have many other undesirable characteristics, for example, high toxicity, other biological activities, insolubility, or metabolism problems. The structure of the lead compound is then modified by synthesis to amplify the desired activity and to minimize or eliminate the unwanted properties. Prior to an elaboration of approaches to lead discovery and lead modification, two of the rare drugs discovered without a lead are discussed.

A. Drug Discovery without a Lead

1. Penicillins

In 1928 Alexander Fleming noticed a green mold growing in a culture of *Staphylococcus aureus*, and where the two had converged, the bacteria were lysed.[1] This led to the discovery of penicillin, which was produced by the mold. It may be thought that this observation was made by other scientists who just ignored it, and, therefore, Fleming was unique for following up on it. However, this is not the case. Fleming tried many times to rediscover this phenomenon without success; it was his colleague, Dr. Ronald Hare,[2,3] who was able to reproduce the observation. It only occurred the first time because a combination of unlikely events all took place simultaneously. Hare found that very special conditions were required to produce the phenomenon initially observed by Fleming. The culture dish inoculated by Fleming must have become accidentally and simultaneously contaminated with the mold spore. Instead of placing the dish in the refrigerator or incubator when he went on vacation as is normally done, Fleming inadvertently left it on his lab bench. When he returned the following month, he noticed the lysed bacteria. Ordinarily, penicillin does not lyse these bacteria; it prevents them from developing, but it has no effect if added after the bacteria have developed. However, while Fleming was on vacation (July to August) the weather was unseasonably cold, and this provided the particular temperature required for the mold and the staphylococci to grow slowly and produce the lysis. Another extraordinary circumstance was that the particular strain of the mold on Fleming's culture was a relatively good penicillin producer, although most strains of that mold (*Penicillium*) produce no penicillin at all. The mold presumably came from the laboratory just below Fleming's where research on molds was going on at the time.

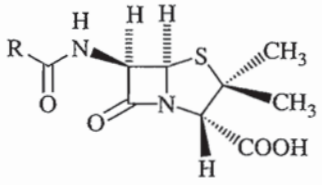
Although Fleming suggested that penicillin could be useful as a topical antiseptic, he was not successful in producing penicillin in a form suitable to treat infections. Nothing more was done until Sir Howard Florey at Oxford University reinvestigated the possibility of producing penicillin in a useful form. In 1940 he succeeded in producing penicillin that could be administered topically and systemically,[4] but the full extent of the value of penicillin was not revealed until the late 1940s.[5] Two reasons for the delay in the universal utilization of penicillin were the emergence of the sulfonamide antibacterials (sulfa drugs, **2.1**; see Chapter 5, Section IV,B,1) in 1935 and the outbreak of World War II. The pharmacology, production, and clinical application of penicillin were not revealed until after the war so that this wonder drug would



2.1

not be used by the Germans. A team of Allied scientists who were interrogating German scientists involved in chemotherapeutic research were told that the Germans thought the initial report of penicillin was made just for commercial reasons to compete with the sulfa drugs. They did not take the report seriously.

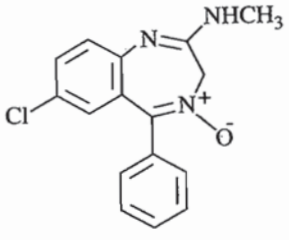
The original mold was *Penicillium notatum*, a strain that gave a relatively low yield of penicillin. It was replaced by *Penicillium chrysogenum*,[6] which had been cultured from a mold growing on a grapefruit in a market in Peoria, Illinois! The correct structure of penicillin (**2.2**) was elucidated in 1943 by Sir Robert Robinson (Oxford) and Karl Folkers (Merck). Several different penicillin analogs (R group varied) were isolated early on; only two of these (**2.2**, R = $PhOCH_2$, penicillin V, and **2.2**, R = CH_2Ph, penicillin G) are still in use today.



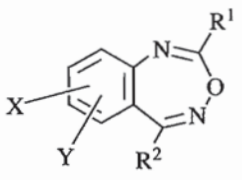
2.2

2. Librium

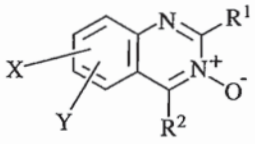
The first benzodiazepine tranquilizer drug, Librium [7-chloro-2-(methylamino)-5-phenyl-3*H*-1,4-benzodiazepine 4-oxide, **2.3**], was discovered serendipitously.[7] Dr. Leo Sternbach at Roche was involved in a program to synthesize a new class of tranquilizer drugs. He originally set out to prepare a series of benzheptoxdiazines (**2.4**), but when R^1 was CH_2NR_2 and R^2 was C_6H_5, it was found that the actual structure was that of a quinazoline 3-oxide (**2.5**). However, none of these compounds gave any interesting pharmacological results. The program was abandoned in 1955 in order for Sternbach to work on a different project. In 1957 during a general laboratory cleanup a vial containing what was thought to be **2.5** (X = 7-Cl, $R^1 = CH_2NHCH_3$, $R^2 = C_6H_5$) was found and, as a last effort, was submitted for pharmacological testing. Unlike all the other compounds submitted, this one gave very promising results in six different tests used for preliminary screening of tranquilizers.



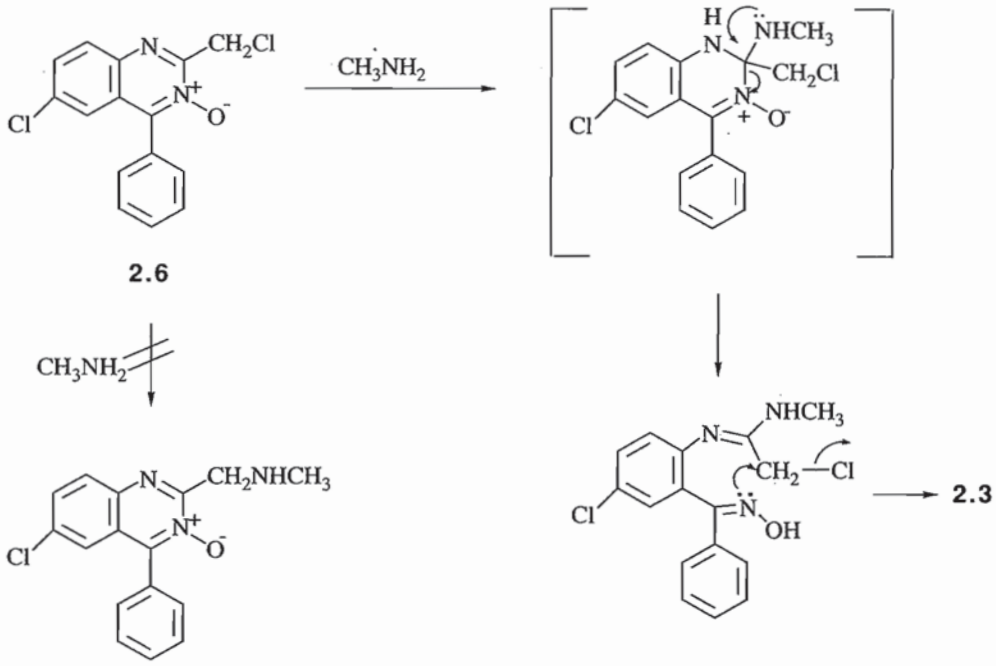
2.3



2.4



2.5



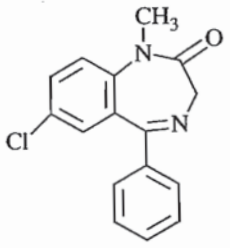
Scheme 2.1. Mechanism for formation of Librium.

Further investigation revealed that the compound was not a quinazoline 3-oxide but, rather, was the benzodiazepine 4-oxide (**2.3**), presumably produced in an unexpected reaction of the corresponding chloromethyl quinazoline 3-oxide (**2.6**) with methylamine (Scheme 2.1). If this compound had not been found in the laboratory cleanup, all of the negative pharmacological results would have been reported for the quinazoline 3-oxide class of compounds, and benzodiazepine 4-oxides may not have been discovered for many years to come.

The examples of drug discovery without a lead are quite few in number. The typical occurrence is that a lead compound is identified and its structure is modified to give, eventually, the drug that goes to the clinic.

B. Lead Discovery

Penicillin V and Librium are, indeed, two important drugs that were discovered without a lead. Once they were identified, however, they then became lead compounds for future analogs. There are now a myriad of penicillin-derived antibacterials that have been synthesized as the result of the structure elucidation of the earliest penicillins. Valium (diazepam, **2.7**) was synthesized at Roche even before Librium was introduced on to the market; this drug was derived from the lead compound Librium and is almost 10 times more potent than the lead.



2.7

In general, the difficulty arises in the discovery of the lead compound. There are several approaches that can be taken to identify a lead. The first requirement for all of the approaches is to have a means to assay compounds for a particular biological activity, so that it will be known when a compound is active. A *bioassay* (or *screen*) is a means of determining in a biological system, relative to a control compound, whether a compound has the desired activity and, if so, what the relative potency of the compound is. Note the distinction between the terms activity and potency. *Activity* is the particular biological or pharmacological effect (e.g., antibacterial activity or anticonvulsant activity); *potency* is the strength of that effect. Some bioassays (or screens) begin as *in vitro* tests, for example, the inhibition of an enzyme or antagonism of a receptor; others are *in vivo* tests, for example, the ability of the compound to prevent an induced seizure in a mouse. In general, the *in vitro* tests are quicker and less expensive. Once the bioassay is developed, there are a variety of approaches to identify a lead.

1. Random Screening

Random screening involves no intellectualization; all compounds are tested in the bioassay without regard to their structures. Prior to 1935 (the discovery of sulfa drugs), random screening was essentially the only approach; today this method is used to a lesser degree. However, random screening programs are still very important in order to discover drugs or leads that have unexpected and unusual structures for various targets.

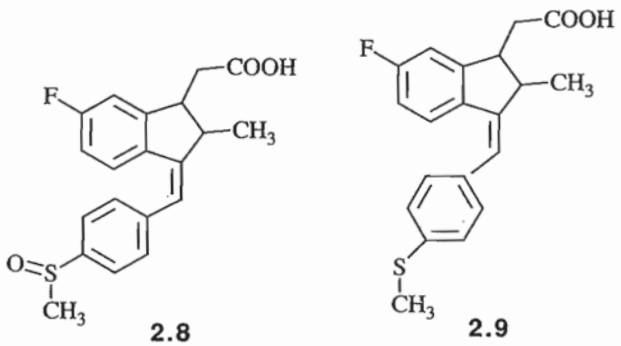
The two major classes of materials screened are synthetic chemicals and natural products (microbial, plant, and marine). An example of a random screen of synthetic and natural compounds is the "war on cancer" declared by Congress and the National Cancer Institute in the early 1970s. Any new compound submitted was screened in a mouse tumor bioassay. Few new anticancer drugs resulted from that screen, but many known anticancer drugs also did not show activity in the screen used. As a result of that observation, multiple bioassay systems are now utilized. In the 1940s and 1950s a random screen by various pharmaceutical companies of soil samples in search of new antibiotics was undertaken. In this case, however, not only were numerous leads uncovered, but two important antibiotics, streptomycin and the tetracyclines, were found.

2. Nonrandom Screening

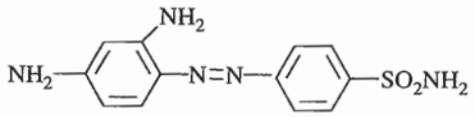
Nonrandom screening is a slightly more narrow approach than is random screening. In this case compounds having a vague resemblance to weakly active compounds uncovered in a random screen or compounds containing different functional groups than leads may be tested selectively. By the late 1970s the National Cancer Institute's random screen was modified to a nonrandom screen because of budgetary and manpower restrictions. Also, the single tumor screen was changed to a variety of tumor screens, as it was realized that cancer is not just a single disease.

3. Drug Metabolism Studies

During metabolism studies *drug metabolites* (drug degradation products generated *in vivo*) that are isolated are screened in order to determine if the activity observed is derived from the drug candidate or from a metabolite. For example, the anti-inflammatory drug sulindac (**2.8**) is not the active agent; the metabolic reduction product, **2.9**, is responsible for the activity.[8] A classic example of this approach is the discovery of the antibacterial agent sulfanilamide (**2.1**, R = H), which was found to be a metabolite of prontosil (**2.10**) (see Chapter 5, Section IV,B,1 for details).



2.8 **2.9**



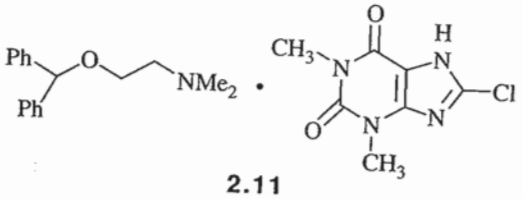
2.10

4. Clinical Observations

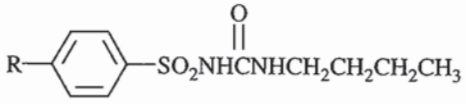
Often a drug candidate during animal testing or clinical trials will exhibit more than one pharmacological activity; that is, it may produce a side effect. This compound, then, can be used as a lead for the secondary activity. In 1947 an antihistamine, dimenhydrinate (**2.11**; Dramamine®) was tested at the allergy clinic at Johns Hopkins University and was found also to be effective in

relieving a patient who suffered from car sickness; a further study proved its effectiveness in the treatment of seasickness[9] and airsickness.[10] It is now the most widely used drug for the treatment of all forms of motion sickness.

An antibacterial agent, carbutamide (**2.12**, R = NH_2), was found to have an antidiabetic side effect. However, it could not be used as an antidiabetic drug because of its antibacterial activity. Carbutamide, then, was a lead for the discovery of tolbutamide (**2.12**, R = CH_3), an antidiabetic drug without antibacterial activity.



2.11

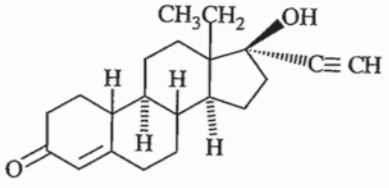


2.12

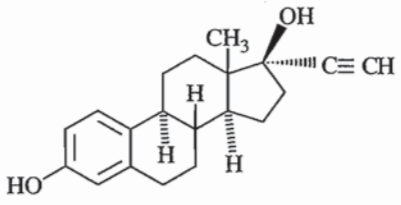
5. Rational Approaches to Lead Discovery

None of the above approaches to lead discovery involves a major rational component. The lead is just found by screening techniques, as a by-product of drug metabolism studies, or from whole animal investigations. Is it possible to design a compound having a particular activity? Rational approaches to drug design have now become the major routes to lead discovery. The first step is to identify the cause for the disease state. Most diseases, or at least the symptoms of diseases, arise from an imbalance of particular chemicals in the body, from the invasion of a foreign organism, or from aberrant cell growth. As discussed in later chapters, the effects of the imbalance can be corrected by antagonism or agonism of a receptor (see Chapter 3) or by inhibition of a particular enzyme (see Chapter 5). Foreign organism enzyme inhibition and interference with DNA biosynthesis or function (see Chapter 6) are also important approaches to treat diseases arising from microorganisms and aberrant cell growth.

Once the relevant biochemical system is identified, lead compounds then become the natural receptor agonists or enzyme substrates. For example, lead compounds for the contraceptives (+)-norgestrel (**2.13**) and 17α-ethynylestradiol (**2.14**) were the steroidal hormones progesterone (**2.15**) and 17β-estradiol

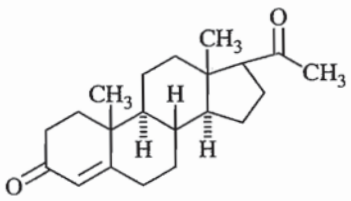


2.13

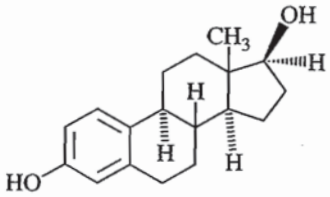


2.14

(**2.16**). Whereas the steroid hormones **2.15** and **2.16** show weak and short-lasting effects, the oral contraceptives **2.13** and **2.14** exert strong progestational activity of long duration.

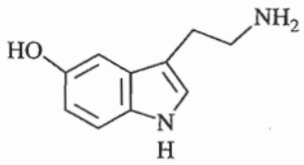


2.15

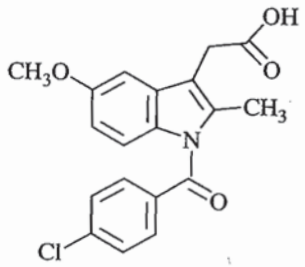


2.16

At Merck it was believed that serotonin (**2.17**) was a possible mediator of inflammation. Consequently, serotonin was used as a lead for anti-inflammatory agents, and from this lead the anti-inflammatory drug indomethacin (**2.18**) was developed.[11]



2.17



2.18

The rational approaches are directed at lead discovery. It is not possible, with much accuracy, to foretell toxicity and side effects, anticipate transport characteristics, or predict the metabolic fate of a drug. Once a lead is identified, its structure can be modified until an effective drug is prepared.

II. Drug Development: Lead Modification

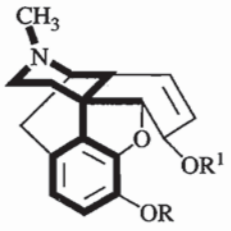
Once your lead compound is in hand, how do you know what to modify in order to improve the desired pharmacological properties?

A. Identification of the Active Part: The Pharmacophore

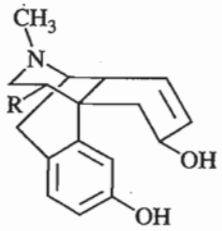
Interactions of drugs with receptors are very specific (see Chapter 3). Therefore, only a small part of the lead compound may be involved in the appropriate interactions. The relevant groups on a molecule that interact with a recep-

tor and are responsible for the activity are collectively known as the *pharmacophore*. If the lead compound has additional groups, they may interfere with the appropriate interactions. One approach to lead modification is to cut away sections of the molecule in order to determine what parts are essential and which are superfluous.

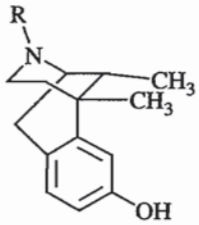
As an example of how a molecule can be trimmed and still result in increased potency or modified activity, consider the addictive analgetics morphine (**2.19**, R = R′ = H), codeine (**2.19**, R = CH_3, R′ = H), and heroin (R = R′ = $COCH_3$). The pharmacophore is darkened. If the dihydrofuran oxygen is excised, morphinan (**2.20**, R = H)[12] results; the hydroxy analog levorphanol[13] (**2.20**, R = OH) is 3 to 4 times more potent than morphine as an analgetic, but it retains the addictive properties. Removal of half of the cyclohexene ring, leaving only methyl substituents, gives benzomorphan (**2.21**, R = CH_3).[14] This compound shows some separation of analgetic and addictive effects; cyclazocine (**2.21**, R = CH_2— ) and pentazocine [**2.21**, R = $CH_2CH{=}C(CH_3)_2$] are analogs with much lower addiction liabilities. Cutting away the cyclohexane fused ring (**2.22**) also has little effect on the analgetic activity in animal tests. Removal of all fused rings, for example, in the case of meperidine (**2.23**, Demerol®), gives an analgetic still possessing 10–12% of the overall potency of morphine.[15] Even acyclic analogs are active. Dextropropoxyphene (**2.24**, Darvon®) is one-half to two-thirds as potent as codeine; its activity can be ascribed to the fact that it can assume a conformation related to that of the morphine pharmacophore. Another acyclic analog is methadone (**2.25**) which is as potent an analgetic as morphine; the (−)-isomer is used in the treatment of opioid abstinence syndromes in heroin abusers.



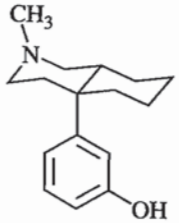
2.19



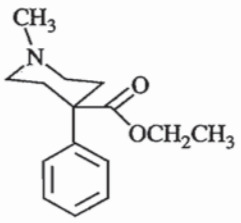
2.20



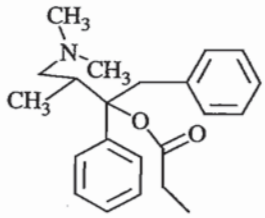
2.21



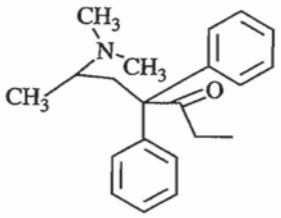
2.22



2.23



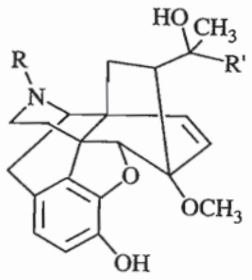
2.24



2.25

In some cases an increase in structural complexity and/or rigidity can lead to increased potency. For example, an oripavine derivative such as etorphine (**2.26**, R = CH_3, R′ = C_3H_7), which has a two-carbon bridge and a substituent not in morphine, is about 1000 times more potent than morphine[16] and, therefore, is used in veterinary medicine to immobilize large animals. The related analog, buprenorphine (**2.26**, R = CH_2—  , R′ = *tert*-Bu, double bond reduced) is 10–20 times more potent than morphine and has a very low level of dependence liability. Apparently, the additional rigidity of the oripavine derivatives increases the appropriate receptor interactions (see Chapter 3).

Once the pharmacophore is identified, manipulation of functional groups becomes consequential.

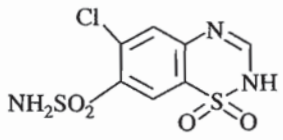


2.26

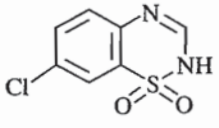
B. Functional Group Modification

The importance of functional group modification was seen in Section I,B,4; the amino group of carbutamide (**2.12**, R = NH_2) was replaced by a methyl group to give tolbutamide (**2.12**, R = CH_3), and in so doing the antibacterial activity was separated away from the antidiabetic activity. In some cases an experienced medicinal chemist knows what functional group will elicit a particular effect. Chlorothiazide (**2.27**) is an antihypertensive agent that has a strong diuretic (increased urine excretion) effect as well. It was known from sulfanilamide work that the sulfonamide side chain can give diuretic activity (see Section II,C). Consequently, diazoxide (**2.28**) was prepared as an antihypertensive drug without diuretic activity.

There, obviously, is a relationship between the molecular structure of a compound and its activity. This phenomenon was first realized over 120 years ago.



2.27



2.28

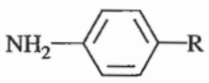
C. Structure–Activity Relationships

In 1868 Crum-Brown and Fraser,[17] suspecting that the quaternary ammonium character of curare may be responsible for its muscular paralytic properties, examined the neuromuscular blocking effects of a variety of simple quaternary ammonium salts and quaternized alkaloids in animals. From these studies they concluded that the physiological action of a molecule was a function of its chemical constitution. Shortly thereafter, Richardson[18] noted that the hypnotic activity of aliphatic alcohols was a function of their molecular weight. These observations are the basis for future structure–activity relationships (SAR).

Drugs can be classified as being structurally specific or structurally nonspecific. *Structurally specific drugs*, which most drugs are, act at specific sites, such as a receptor or an enzyme. Their activity and potency are very susceptible to small changes in chemical structure; molecules with similar biological activities tend to have common structural features. *Structurally nonspecific drugs* have no specific site of action and usually have lower potency. Similar biological activities may occur with a variety of structures. Examples of these drugs are gaseous anesthetics, sedatives and hypnotics, and many antiseptics and disinfectants.

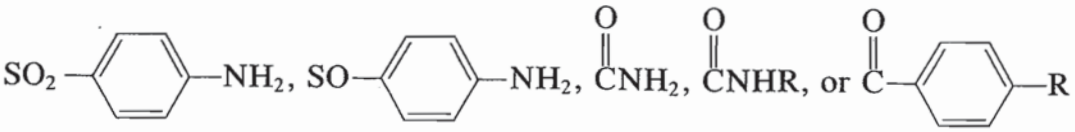
Even though only a part of the molecule may be associated with the activity, there are a multitude of molecular modifications that could be made. Early SAR studies (prior to the 1960s) simply involved the syntheses of as many analogs as possible of the lead and their testing to determine the effect of structure on activity (or potency). Once enough analogs were prepared and sufficient data accumulated, conclusions could be made regarding structure–activity relationships.

An excellent example of this approach came from the development of the sulfonamide antibacterial agents (sulfa drugs). After a number of analogs of the lead compound sulfanilamide (**2.1**, R = H) were prepared, it was found that compounds of this general structure exhibited diuretic and antidiabetic activities as well as antimicrobial activity. Compounds with each type of activity eventually were shown to possess certain structural features in common. On the basis of the biological results of greater than 10,000 compounds, several SAR generalizations have been made.[19] Antimicrobial agents have structure **2.29** (R = SO_2NHR' or SO_3H) where (1) the amino and sulfonyl groups on the benzene ring should be para; (2) the anilino amino group may be unsubstituted (as shown) or may have a substituent that is removed *in vivo*; (3) replacement of the benzene ring by other ring systems, or the introduction of



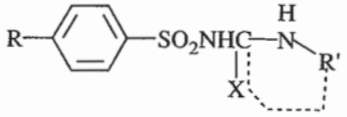
2.29

additional substituents on it, decreases the potency or abolishes the activity; (4) R may be



but the potency is reduced in most cases; (5) N′-monosubstitution (R = SO_2NHR') results in more potent compounds, and the potency increases with heteroaromatic substitution; and (6) N′-disubstitution (R = $SO_2NR'_2$), in general, leads to inactive compounds.

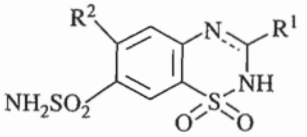
Antidiabetic agents are compounds with structure **2.30**, where X may be O, S, or N incorporated into a heteroaromatic structure such as a thiadiazole or a pyrimidine or in an acyclic structure such as a urea or thiourea. In the case of ureas, the N^2 should carry as a substituent a chain of at least two carbon atoms.[20]



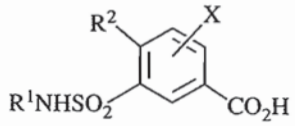
2.30

Sulfonamide diuretics are of two general structural types, hydrochlorothiazides (**2.31**) and the high ceiling type (**2.32**). The former compounds have 1,3-disulfamyl groups on the benzene ring, and R^2 is an electronegative group such as Cl, CF_3, or NHR. The high ceiling compounds contain 1-sulfamyl-3-carboxy groups. Substituent R^2 is Cl, Ph, or PhZ, where Z may be O, S, CO, or NH, and X can be at position 2 or 3 and is normally NHR, OR, or SR.[21]

The sulfonamide example is strong evidence to support the notion that a correlation does exist between structure and activity, but how do you know what molecular modifications to make in order to fine-tune the lead compound?



2.31



2.32

D. Structure Modifications to Increase Potency and Therapeutic Index

In the preceding section it was made clear that structure modifications were the keys to activity and potency manipulations. After years of structure–activity relationship studies, various standard molecular modification ap-

proaches have been developed for the systematic improvement of the *therapeutic index* (also called the *therapeutic ratio*), which is a measure of the ratio of undesirable to desirable drug effects. For *in vivo* systems the therapeutic index could be the ratio of the LD_{50} (the lethal dose for 50% of the test animals) to the ED_{50} (the effective dose that produces the maximum therapeutic effect in 50% of the test animals). The larger the therapeutic index, the greater the margin of safety of the compound. A number of these structural modification methodologies follow.

1. Homologation

A *homologous series* is a group of compounds that differ by a constant unit, generally a CH_2 group. As will become more apparent in Section II,E, biological properties of homologous compounds show regularities of increase and decrease. For many series of compounds, lengthening of a saturated carbon side chain from one (methyl) to five to nine atoms (pentyl to nonyl) produces an increase in pharmacological effects; further lengthening results in a sudden decrease in potency (Fig. 2.1). In Section II,E,2,b it will be shown that this phenomenon corresponds to increased lipophilicity of the molecule, which permits penetration into cell membranes' until its lowered water solubility becomes problematic in its transport through aqueous media. In the case of aliphatic amines another problem is micelle formation, which begins at about C_{12}. This effectively removes the compound from potential interaction with the appropriate receptors. One of, if not the, earliest example of this potency versus chain length phenomenon was reported by Richardson,[18] who was

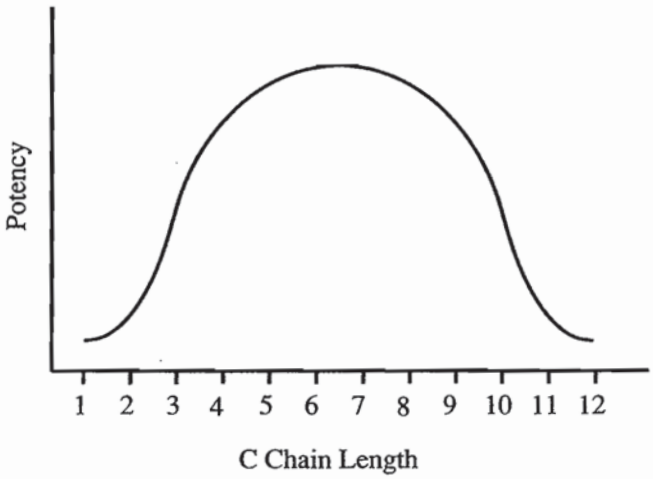
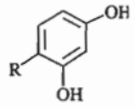
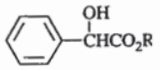


Figure 2.1. General effect of carbon chain length on drug potency.

Table 2.1 Effect of Chain Length on Potency: Antibacterial Activity of 4-*n*-Alkylresorcinols[22a] and Spasmolytic Activity of Mandelate Esters[22b]





R	Phenol coefficient	% Spasmolytic activity[a]
methyl	—	0.3
ethyl	—	0.7
n-propyl	5	2.4
n-butyl	22	9.8
n-pentyl	33	28
n-hexyl	51	35
n-heptyl	30	51
n-octyl	0	130
n-nonyl	0	190
n-decyl	0	37
n-undecyl	0	22
i-propyl	—	0.9
i-butyl	15.2	8.3
i-amyl	23.8	28
i-hexyl	27	—

[a] Relative to 3,3,5-trimethylcyclohexanol, set at 100%.

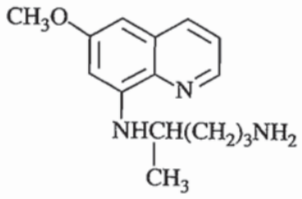
investigating the hypnotic activity of alcohols. The maximum effect occurred for 1-hexanol to 1-octanol; then the potency declined upon chain lengthening until no activity was observed for hexadecanol.

A study by Dohme *et al.*[22a] on 4-alkyl-substituted resorcinol derivatives showed that the peak antibacterial activity occurred with 4-*n*-hexylresorcinol (see Table 2.1), a compound now used as a topical anesthetic in a variety of throat lozenges. Funcke *et al.*[22b] found that the peak spasmolytic activity of a series of mandelate esters occurred with the *n*-nonyl ester (see Table 2.1).

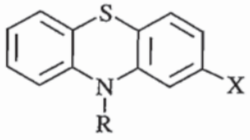
2. Chain Branching

When a simple lipophilic relationship is important as described above, then *chain branching* lowers the potency of a compound. This phenomenon is exemplified by the lower potency of the compounds having isoalkyl chains in Table 2.1. Chain branching also can interfere with receptor binding. For example, phenethylamine ($PhCH_2CH_2NH_2$) is an excellent substrate for monoamine oxidase [amine oxidase (flavin-containing)], but α-methylphenethylamine (amphetamine) is a poor substrate. Primary amines often are more potent than secondary amines which are more potent than tertiary amines. For example, the antimalarial drug primaquine (**2.33**) is much more potent than its secondary or tertiary amine homologs.

Major pharmacological changes can occur with chain branching and homologation. Consider the 10-aminoalkylphenothiazines (**2.34**, X = H). When R is $CH_2CH(CH_3)N(CH_3)_2$ (promethazine) or $CH_2CH_2N(CH_3)_2$ (diethazine), antispasmodic and antihistaminic activities predominate. However, the homolog **2.34** with R being $CH_2CH_2CH_2N(CH_3)_2$ (promazine) has greatly reduced antispasmodic and antihistaminic activities, but sedative and tranquilizing activities are greatly enhanced. In the case of the branched chain analog **2.34** with R equal to $CH_2CH(CH_3)CH_2N(CH_3)_2$ (trimeprazine), the tranquilizing activity is reduced and antipruritic (anti-itch) activity increases.



2.33



2.34

3. Ring–Chain Transformations

Another modification that can be made is the transformation of alkyl substituents into cyclic analogs. Consider the promazines again (**2.34**). Chlorpromazine [**2.34**, X = Cl, R = $CH_2CH_2CH_2N(CH_3)_2$] and **2.34** (X = Cl, R = $CH_2CH_2CH_2N$⟨cyclopentane⟩) are equivalent as tranquilizers in animal tests. Trimeprazine [**2.34**, X = H, R = $CH_2CH(CH_3)CH_2N(CH_3)_2$] and methdilazine [**2.34**, X = H, R = CH_2—CH(—CH_2—CH_2—)—N—CH_3] have similar antipruritic activity in man.

Different activities can result from a ring–chain transformation as well. For example, if the dimethylamino group of chlorpromazine is substituted by a methylpiperazine ring (**2.34**, X = Cl, R = $CH_2CH_2CH_2N$⟨piperazine⟩NCH_3; prochlorperazine), the antiemetic (prevents nausea and vomiting) activity is greatly enhanced. In this case, however, an additional amino group is added.

4. Bioisosterism

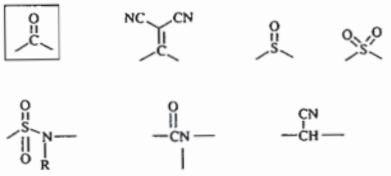
Bioisosteres are substituents or groups that have chemical or physical similarities, and which produce broadly similar biological properties.[23] Bioisosterism is a lead modification approach that has been shown to be useful to attenuate toxicity or to modify the activity of a lead, and it may have a significant role in the alteration of metabolism of a lead. There are classical isosteres[24,25] and nonclassical isosteres.[23,26] In 1925 Grimm[27] formulated the *hydride displacement law* to describe similarities between groups that have the same number of valence electrons but may have a different number of atoms. Erlenmeyer[28] later redefined isosteres as atoms, ions, or molecules in which the peripheral layers of electrons can be considered to be identical. These two definitions describe *classical isosteres*; examples are shown in Table 2.2. *Nonclassical*

Table 2.2 Classical Isosteres[24,25]

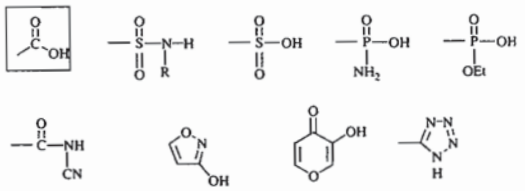
1. Univalent atoms and groups					
a. CH_3	NH_2	OH	F	Cl	
b. Cl	PH_2	SH			
c. Br	*i*-Pr				
d. I	*t*-Bu				
2. Bivalent atoms and groups					
a. $-CH_2-$	—NH—	—O—	—S—	—Se—	
b. $-COCH_2R$	—CONHR	$-CO_2R$	—COSR		
3. Trivalent atoms and groups					
a. —CH=	—N=				
b. —P=	—As=				
4. Tetravalent atoms					
a. $-\overset{|}{\underset{|}{C}}-$	$-\overset{|}{\underset{|}{Si}}-$				
b. =C=	$=\overset{+}{N}=$	$=\overset{+}{P}=$			
5. Ring equivalents					
a. —CH=CH—	—S—	(e.g., benzene, thiophene)			
b. —CH=	—N=	(e.g., benzene, pyridine)			
c. —O—	—S—	$-CH_2-$	—NH—	(e.g., tetrahydrofuran, tetrahydrothiophene, cyclopentane, pyrrolidine)	

Table 2.3 Nonclassical Bioisosteres[23]

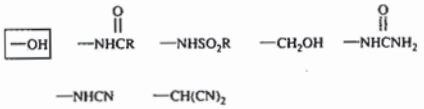
1. **Carbonyl group**



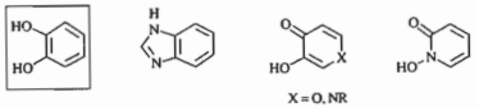
2. **Carboxylic acid group**



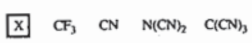
3. **Hydroxy group**



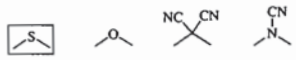
4. **Catechol**



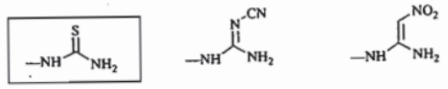
5. **Halogen**



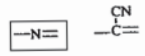
6. **Thioether**



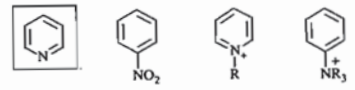
7. **Thiourea**



8. **Azomethine**



9. **Pyridine**

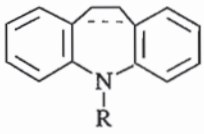


10. **Spacer group**

11. **Hydrogen**

bioisosteres do not have the same number of atoms and do not fit the steric and electronic rules of the classical isosteres, but they do produce a similarity in biological activity. Examples of these are shown in Table 2.3.

Ring–chain transformations also can be considered to be isosteric interchanges. There are hundreds of examples of compounds that differ by a bioisosteric interchange[23,26]; some examples are shown in Table 2.4. Bioisosterism also can lead to changes in activity. If the sulfur atom of the phenothiazine neuroleptic drugs (**2.34**) is replaced by the —CH=CH— or —CH_2CH_2— bioisosteres, then dibenzazepine antidepressant drugs (**2.35**) result.



2.35

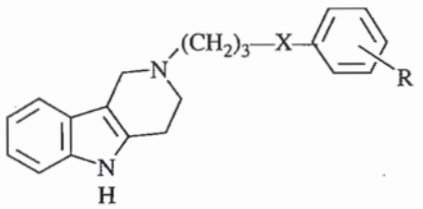
It is, actually, quite surprising that bioisosterism should be such a successful approach to lead modification. Perusal of Table 2.2, and especially of Table 2.3, makes it clear that in making a bioisosteric replacement, one or more of the following parameters will change: size, shape, electronic distribution, lipid solubility, water solubility, pK_a, chemical reactivity, and hydrogen bonding. Because a drug must get to the site of action, then interact with it (see Chapter 3), modifications made to a molecule may have one or more of the following effects:

1. Structural. If the moiety that is replaced by a bioisostere has a structural role in holding other functionalities in a particular geometry, then size, shape, and hydrogen bonding will be important.
2. Receptor interactions. If the moiety replaced is involved in a specific interaction with a receptor or enzyme, then all of the parameters except lipid and water solubility will be important.
3. Pharmacokinetics. If the moiety replaced is necessary for absorption, transport, and excretion (collectively, with metabolism, termed *pharmacokinetics*) of the compound, then lipophilicity, hydrophilicity, pK_a, and hydrogen bonding will be important.
4. Metabolism. If the moiety replaced is involved in blocking or aiding metabolism, then the chemical reactivity will be important.

It is because of these subtle changes that bioisosterism is effective. This approach allows the medicinal chemist to tinker with only some of the parameters in order to augment the potency, selectivity, and duration of action and to reduce toxicity. Multiple alterations may be necessary to counterbalance effects. For example, if modification of a functionality involved in binding also decreases the lipophilicity of the molecule, thereby reducing its ability to

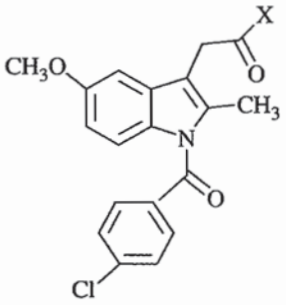
Table 2.4 Examples of Bioisosteric Analogs[23,26]

1. **Neuroleptics (antipsychotics)**



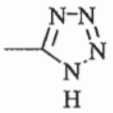
X=  or CHCN

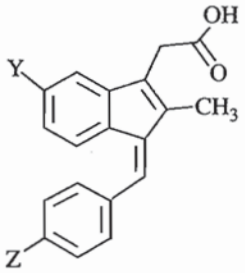
2. **Anti-inflammatory agents**



X = OH (indomethacin)

= NHOH

= 



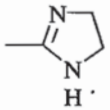
Y = CH_3O Z = Cl

Y = F Z = SCH_3 (sulindac)

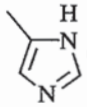
3. **Antihistamines**

R — X — $(CH_2)_n$ — Y X = NH, O, CH_2

Y = $N(CH_3)_2$ (n = 2)



(n = 1)



(n = 1, 2)

penetrate cell walls and cross other membranes, the molecule can be substituted with additional lipophilic groups at sites distant from that involved with binding. Modifications of this sort may change the overall molecular shape and result in another activity.

Up to this point we have been discussing more or less random molecular modifications to make qualitative differences in a lead. In 1868 Crum-Brown and Fraser[17] predicted that some day a mathematical relationship between structure and activity would be expressed. It was not until almost 100 years later that this prediction began to be realized and a new era in drug design was born. In 1962 Corwin Hansch attempted to quantify the effects of particular substituent modifications, and from this quantitative structure–activity relationship (QSAR) studies developed.[29]

E. Quantitative Structure–Activity Relationships

1. Historical

The concept of quantitative drug design is based on the fact that the biological properties of a compound are a function of its *physicochemical parameters*, that is, physical properties, such as solubility, lipophilicity, electronic effects, ionization, and stereochemistry, that have a profound influence on the chemistry of the compounds. The first attempt to relate a physicochemical parameter to a pharmacological effect was reported in 1893 by Richet.[30] He observed that the narcotic action of a group of organic compounds was inversely related to their water solubility (Richet's rule). Overton[31] and Meyer[32] related tadpole narcosis induced by a series of nonionized compounds added to the water in which the tadpoles were swimming to the ability of the compounds to partition between oil and water. These early observations regarding the depressant action of structurally nonspecific drugs were rationalized by Ferguson.[33] He reasoned that, for a state of equilibrium, simple thermodynamic principles could be applied to drug activities, and that the important parameter for correlation of narcotic activities was the relative saturation (termed *thermodynamic activity* by Ferguson) of the drug in the external phase or extracellular fluids. This is known as *Ferguson's principle*, which is useful for the classification of the general mode of action of a drug and for predicting the degree of its biological effect. The numerical range of the thermodynamic activity for structurally nonspecific drugs is 0.01 to 1.0, indicating that they are active only at relatively high concentrations. Structurally specific drugs have thermodynamic activities considerably less than 0.01 and normally below 0.001.

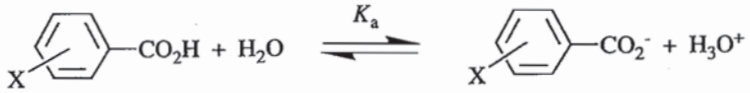
In 1951 Hansch *et al.*[34] noted a correlation between the plant growth activity of phenoxyacetic acid derivatives and the electron density at the ortho position (lower electron density gave increased activity). They made an at-

tempt to quantify this relationship by the application of the Hammett σ functions (see Section II,E,2,a), but this was unsuccessful.

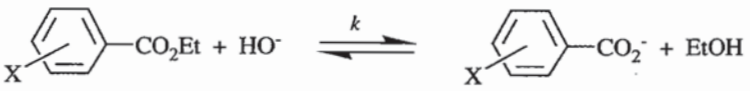
2. Physicochemical Parameters

a. Electronic Effects: The Hammett Equation. In 1940 L. P. Hammett published a book entitled *Physical Organic Chemistry*[35] that marked the beginning of a new era in organic chemistry, namely, quantitative organic chemistry. Hammett's postulate was that the electronic effects (both the inductive and resonance effects) of a set of substituents on different organic reactions should be similar. Therefore, if numbers could be assigned to substituents in a standard organic reaction, these same numbers could be used to estimate rates in a new organic reaction. This was the first approach toward the prediction of reaction rates. Hammett chose benzoic acids as the standard system.

Consider the reaction shown in Scheme 2.2. Intuitively, it seems reasonable that as X becomes electron withdrawing (relative to H), the equilibrium constant (K_a) should increase (the reaction to the right is favored) because X is inductively pulling electron density from the carboxylic acid group, making it more acidic (ground state argument); it also is stabilizing the incipient negative charge on the carboxylate group in the transition state (transition state argument). Conversely, when X is electron donating, the equilibrium constant should decrease. A similar relationship should exist for a rate constant (k) where charge develops in the transition state. Hammett chose the reaction shown in Scheme 2.3 as the standard system. If K_a is measured from Scheme 2.2 and k from Scheme 2.3 for a series of substituents X, and the data are expressed in a double logarithm plot (Fig. 2.2[36]), then a straight line can be drawn through most of the data points. This is known as a *linear free-energy relationship*. When X is a meta or para substituent, then virtually all of the points fall on the straight line; the ortho substituent points are badly scattered. The Hammett relationship does not hold for ortho substituents because of steric interactions and polar effects.



Scheme 2.2. Ionization of substituted benzoates.



Scheme 2.3. Saponification of substituted ethyl benzoates.

The linear correlation for the meta and para substituents is observed for rate or equilibrium constants for a wide variety of organic reactions. The straight

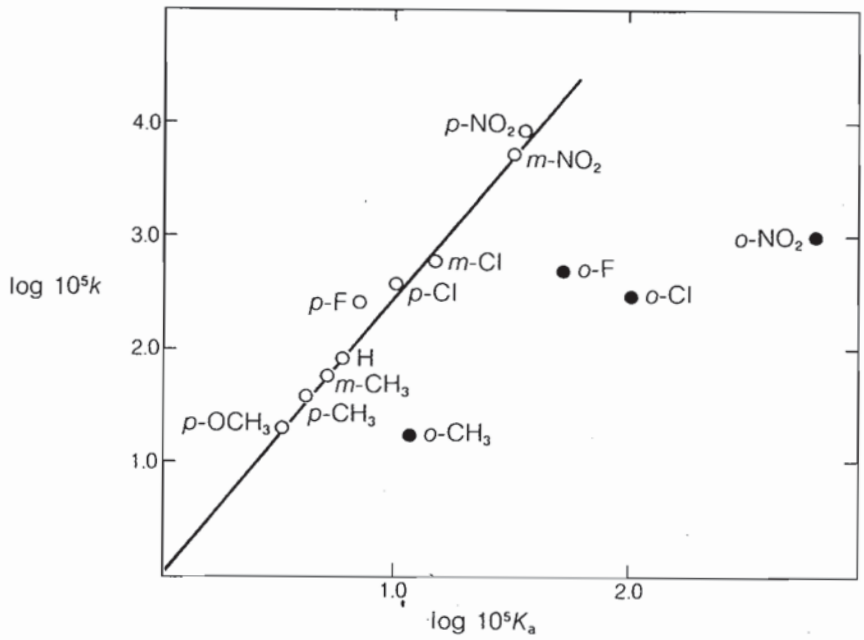


Figure 2.2. Linear free energy relationship for the dissociation of substituted benzoic acids in water at 25°C (K_a) against the rates of alkaline hydrolysis of substituted ethyl benzoates in 85% ethanol–water at 30°C (k). [Reprinted with permission from Roberts, J. D., and Caserio, M. C. (1977). "Basic Principles of Organic Chemistry," 2nd ed., p. 1331. W. A. Benjamin, Menlo Park, CA. Copyright © 1977 Benjamin/Cummings Publishing Company.]

line can be expressed by Eq. (2.1), where the two variables are log k and log K. The slope of the line is ρ and the intercept is C. When there is no substituent, that is, when X is H, then Eq. (2.2) holds. Subtraction of Eq. (2.2) from Eq. (2.1) gives Eq. (2.3), where k and K are the rate and equilibrium constants, respectively, for compounds with a substituent X, and k_o and K_o are the same for the parent compound (X = H). If log K/K_o is defined as σ, then Eq. (2.3) reduces to Eq. (2.4), the *Hammett equation*. The *electronic parameter*, σ, depends on the electronic properties and position of the substituent on the ring and, therefore, is also called the *substituent constant*. The more electron withdrawing a substituent, the more positive is its σ value (relative to H, which is set at 0.0); conversely, the more electron donating, the more negative is its σ value. The meta σ constants result from inductive effects, but the para σ constants correspond to the net inductive and resonance effects. Therefore, σ_{meta} and σ_{para} for the same substituent, generally, are not the same.

$$\log k = \rho \log K + C \tag{2.1}$$

$$\log k_o = \rho \log K_o + C \tag{2.2}$$

$$\log \frac{k}{k_o} = \rho \log \frac{K}{K_o} \tag{2.3}$$

$$\log \frac{k}{k_o} = \rho\sigma \tag{2.4}$$

The ρ values (the slope) depend on the particular type of reaction and the reaction conditions (e.g., temperature and solvent) and, therefore, are called *reaction constants*. The importance of ρ is that it is a measure of the sensitivity of the reaction to the electronic effects of the meta and para substituents. A large ρ, either positive or negative, indicates great sensitivity to substituent effects. Reactions that are favored by high electron density in the transition state have negative ρ values; reactions that are aided by electron withdrawal have positive ρ values.

As we shall see in Section II,E,3, the substituent constant σ will be of major significance to QSAR.

b. Lipophilicity Effects: The Basis for the Hansch Equation. The crucial breakthrough in QSAR came when Hansch and co-workers[29,37] conceptualized the action of a drug as depending on two processes. The first process is the journey of the drug from its point of entry into the body to the site of action (*pharmacokinetics*), and the second process is the interaction of the drug with the specific site (*pharmacodynamics*). Hansch proposed that the first step in the overall process was a random walk, a diffusion process, in which the drug made its way from a dilute solution outside of the cell to a particular site in the cell. This was visualized as being a relatively slow process, the rate of which being highly dependent on the molecular structure of the drug.

For the drug to reach the site of action, it must be able to interact with two different environments, lipophilic (e.g., membranes) and aqueous (the exobiophase, such as the cytoplasm). The cytoplasm of a cell is essentially a dilute solution of salts in water; all living cells are surrounded by a nonaqueous phase, the membrane. The functions of membranes are to protect the cell from water-soluble substances, to form a surface to which enzymes and other proteins can attach in order to produce a localization and structural organization, and to separate solutions of different electrochemical potentials (e.g., in nerve conduction). One of the most important membranes is known as the *blood–brain barrier*, a membrane that surrounds the capillaries of the circulatory system in the brain and protects it from passive diffusion of undesirable chemicals from the bloodstream. This is an important prophylactic boundary, but it also can block the delivery of central nervous system drugs to their site of action.

Although the structure of membranes has not been resolved, the most widely accepted model is the *fluid mosaic model* (Fig. 2.3).[38] In this depiction

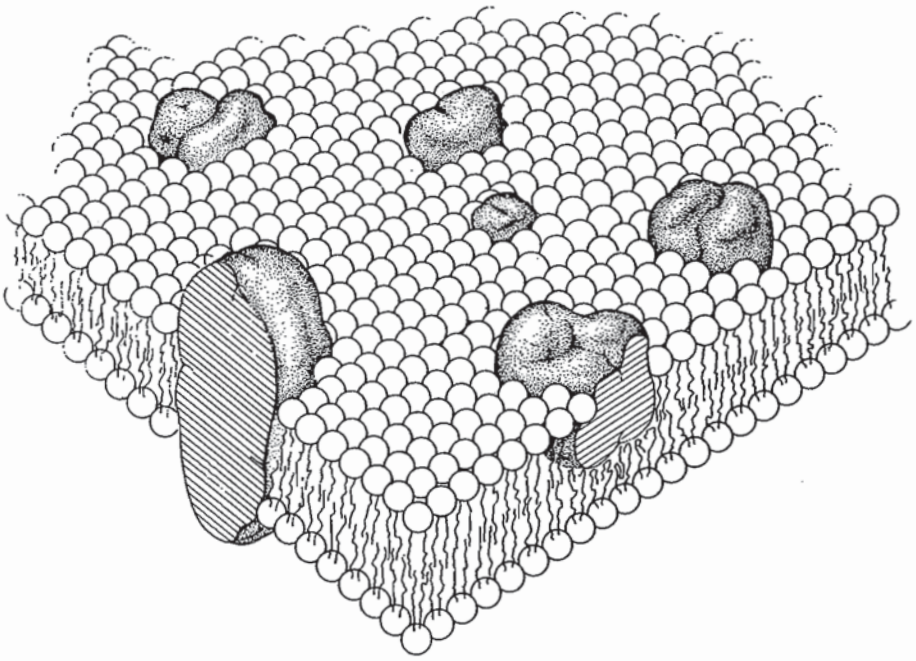
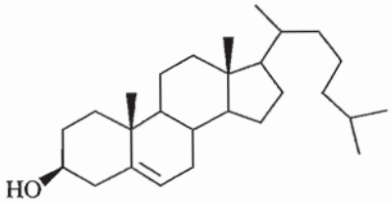
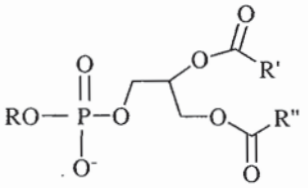


Figure 2.3. Fluid mosaic model of a membrane. The balls represent polar end groups, and the wavy lines are the hydrocarbon chains of the lipids. [From Singer, S. J., and Nicolson, G. L. (1972). *Science* **175,** 720. Copyright © 1972 by the AAAS with permission.]

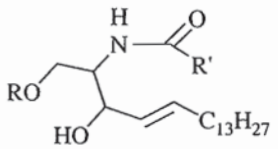
integral proteins are embedded in a lipid bilayer; *peripheral proteins* are associated with only one membrane surface. The structure of the membrane is primarily determined by the structure of the lipids of which it is comprised. The principal classes of lipids found in membranes are neutral cholesterol (**2.36**) and the ionic phospholipids, for example, phosphatidylcholine [**2.37**, R = $(CH_3)_3\overset{+}{N}CH_2CH_2$], phosphatidylethanolamine (**2.37**, R = $\overset{+}{N}H_3CH_2CH_2$), phosphatidylserine [**2.37**, R = $^-OOC(\overset{+}{N}H_3)CHCH_2$), phosphatidylinositol (**2.37**, R = inositol), and sphingomyelin [**2.38**, R = $(CH_3)_3\overset{+}{N}CH_2CH_2O_2PO^-$]; R′CO and R″CO in **2.37** and **2.38** are derived from fatty acids. Glycolipids (**2.38**, R = sugar) also are important membrane constituents.



2.36



2.37



2.38

All of these lipids are *amphipathic*, meaning that one end of the molecule is *hydrophilic* (water soluble) and the other is *hydrophobic* or, if you wish, *lipophilic* (water insoluble; soluble in organic solvents). Thus, the hydroxyl

group in cholesterol, the ammonium groups in the phospholipids, and the sugar residue in the glycolipids are the polar, hydrophilic ends, and the steroid and hydrocarbon moieties are the lipophilic ends. The hydrocarbon part (R′ and R″) actually can be a mixture of chains from 14 to 24 carbon atoms long; approximately 50% of the chains contain a double bond. The polar groups of the lipid bilayer are in contact with the aqueous phase; the hydrocarbon chains project toward each other in the interior with a space between the layers. The stability of the membrane arises from the stabilization of the ionic charges by ion–dipole interactions with the water (see Chapter 3) and from association of the nonpolar groups. The hydrocarbon chains are relatively free to move; therefore, the core is similar to a liquid hydrocarbon.

It occurred to Hansch that the fluidity of the hydrocarbon region of the membrane may explain the correlation noted by Richet,[30] Overton,[31] and Meyer[32] between lipid solubility and biological activity. The Hansch group[29,39] suggested that a reasonable model for the first step in drug action (transport to the site of action) would be the ability of a compound to partition between 1-octanol, which would simulate a lipid membrane, and water (the aqueous phase). 1-Octanol has a long saturated alkyl chain and a hydroxyl group for hydrogen bonding, and it dissolves water to the extent of 1.7 *M* (saturation). This combination of lipophilic chains, hydrophilic groups, and water molecules gives 1-octanol properties very close to those of natural membranes and macromolecules.

Hansch believed that, just as in the case of the Hammett equation, there should be a linear free energy relationship between lipophilicity and biological activity. As a suitable measure of lipophilicity, the *partition coefficient*, *P*, between 1-octanol and water was proposed,[29,39] and *P* was determined by Eq. (2.5), where α is the degree of dissociation of the compound in water calculated from ionization constants.

$$P = \frac{[\text{compound}]_{\text{oct}}}{[\text{compound}]_{\text{aq}}(1 - \alpha)} \tag{2.5}$$

The partition coefficient is derived experimentally by placing a compound in a shaking device (such as a separatory funnel) with varying volumes of 1-octanol and water, determining the concentration of the compound in each layer after mixing, and utilizing Eq. (2.5) to calculate *P*. The value of *P* varies slightly with temperature and concentration of the solute, but with neutral molecules in dilute solutions (<0.01 *M*) and small temperature changes (±5°C), variations in *P* are minor.

Collander[40] had shown previously that the rate of movement of a variety of organic compounds through cellular material was approximately proportional to the logarithm of their partition coefficients between an organic solvent and water. Therefore, as a model for a drug traversing through the body to its site of action, the relative potency of the drug, expressed as log $1/C$, where C is

the concentration of the drug that produces some standard biological effect, was related to its lipophilicity by the parabolic expression shown in Eq. (2.6).[41]

$$\log 1/C = -k(\log P)^2 + k'(\log P) + k'' \qquad (2.6)$$

On the basis of Eq. (2.5), it is apparent that if a compound is more soluble in water than in 1-octanol, P is less than 1, and, therefore, log P is negative. Conversely, a molecule more soluble in 1-octanol has a P value greater than 1, and log P is positive. The larger the value of P, the more there will be an interaction of the drug with the lipid phase (i.e., membranes). As P approaches infinity, the drug interaction will become so great that the drug will not be able to cross the aqueous phase, and it will localize in the first lipophilic phase with which it comes into contact. As P approaches zero, the drug will be so water soluble that it will not be capable of crossing the lipid phase and will localize in the aqueous phase. Somewhere between $P = 0$ and $P = \infty$, there will be a value of P such that drugs having this value will be least hindered in their journey through macromolecules to their site of action. This value is called log P_o, the optimum partition coefficient for biological activity.

This random walk analysis supports the parabolic relationship [Eq. (2.6)] between potency (log $1/C$) and log P (Fig. 2.4). Note the correlation of Fig. 2.4 with the generalization regarding homologous series of compounds (Section II,D,1; Fig. 2.1). An increase in the alkyl chain length increases the lipophilicity of the molecule; apparently, the log P_o generally occurs in the range of 5–9 carbon atoms. Hansch *et al.*[41] found that a number of series of nonspecific hypnotics had similar log P_o values, approximately 2, and they suggested that this is the value of log P_o needed for penetration into the central nervous system (CNS). If a hypnotic agent has a log P considerably different from 2, then its activity probably is derived from mechanisms other than just lipid

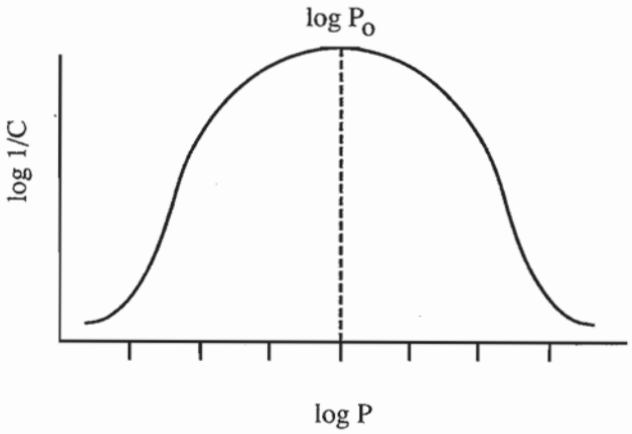


Figure 2.4. Effect of log P on biological response. P is the partition coefficient, and C is the concentration of the compound required to produce a standard biological effect.

transport. If a lead compound has modest CNS activity and has a log P value of 0, it would be reasonable to synthesize an analog with a higher log P.

Can you predict what analog will have a higher log P? In the same way that substituent constants were derived by Hammett for the electronic effects of atoms and groups (σ constants), Hansch and co-workers[29,37,39] derived substituent constants for the contribution of individual atoms and groups to the partition coefficient. The *lipophilicity substituent constant*, π, is defined by Eq. (2.7), which has the same derivation as the Hammett equation. The term P_X is the partition coefficient for the compound with substituent X, and P_H is the partition coefficient for the parent molecule (X = H). As in the case of the Hammett substituent constant σ, π is additive and constitutive. *Additive* means that multiple substituents exert an influence equal to the sum of the individual substituents. *Constitutive* indicates that the effect of a substituent may differ depending on the molecule to which it is attached or on its environment. Alkyl groups are some of the least constitutive. For example, methyl groups attached at the meta or para positions of 15 different benzene derivatives had π_{CH_3} values with a mean and standard derivation of 0.50 ± 0.04. Because of the additive nature of π values, π_{CH_2} can be determined as shown in Eq. (2.8), where the log P values are obtained from standard tables.[42] Because, by definition, $\pi_H = 0$, then $\pi_{CH_2} = \pi_{CH_3}$.

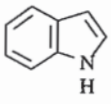
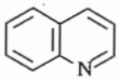
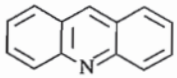
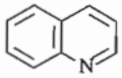
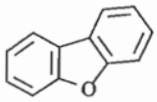
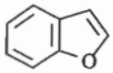
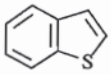
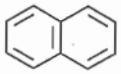
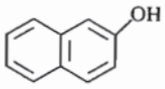
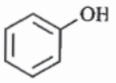
$$\pi = \log P_X - \log P_H = \log \frac{P_X}{P_H} \tag{2.7}$$

$$\begin{aligned} \pi_{CH_2} &= \log P_{\text{nitroethane}} - \log P_{\text{nitromethane}} \\ &= 0.18 - (-0.33) = 0.51 \end{aligned} \tag{2.8}$$

As was alluded to in Section II,D,2 on molecular modification, branching in an alkyl chain lowers the log P or π as a result of the larger molar volumes and shapes of branched compounds. As a rule of thumb, the value of log P or π is lowered by 0.2 unit per branch. For example, the $\pi_{i\text{-Pr}}$ value in 3-isopropylphenoxyacetic acid is 1.30; π_{Pr} is 3(0.5) = 1.50. Another case where π values are fairly constant is conjugated systems, as exemplified by $\pi_{CH=CHCH=CH}$ in Table 2.5.

Inductive effects are quite important to lipophilicity.[43] In general, electron-withdrawing groups increase π when a hydrogen-bonding group is involved. For example π_{CH_2OH} varies as a function of the proximity of an electron-withdrawing phenyl group [Eq. (2.9)],[44] and π_{NO_2} varies as a function of the inductive effect of the nitro group on the hydroxyl group [Eq. (2.10)].[43] The electron-withdrawing inductive effects of the phenyl group [Eq. (2.9)] and the nitro group [Eq. (2.10)] make the nonbonded electrons on the hydroxyl group less available for hydrogen bonding, thereby reducing the affinity of this functional group for the aqueous phase. This, then, increases the log P or π. Also

Table 2.5 Constancy of π for —CH=CH—CH=CH—[41,43]

Log P Difference					$\pi_{CH=CHCH=CH}$
log P 	— log P 	=	2.14 — 0.75 =	1.39	
log P 	— log P 	=	2.03 — 0.65 =	1.38	
log P 	— log P 	=	3.40 — 2.03 =	1.37	
log P 	— log P 	=	4.12 — 2.67 =	1.45	
log P 	— log P 	=	3.12 — 1.81 =	1.31	
log P 	— log P 	=	3.45 — 2.13 =	1.32	
2/3 log P 		=	2/3 (2.13) =	1.42	
log P 	— log P 	=	2.84 — 1.46 =	1.38	

ave. 1.38 ± 0.046

note in Eqs. (2.9) and (2.10) that, because $\pi_H = 0$ by definition, $\log P_{benzene} = \pi_{Ph}$.

$$\pi_{CH_2OH} = \log P_{Ph(CH_2)_2OH} - \log P_{PhCH_3} = -1.33 \quad (2.9)$$
$$\pi_{CH_2OH} = \log P_{PhCH_2OH} - \log P_{PhH} = -1.03$$

$$\pi_{NO_2} = \log P_{pHNO_2} - \log P_{PhH} = -0.28 \quad (2.10)$$
$$\pi_{NO_2} = \log P_{4\text{-}NO_2PhCH_2OH} - \log P_{PhCH_2OH} = 0.11$$

Resonance effects also are important to the lipophilicity much the same way as are inductive effects.[43] Delocalization of nonbonded electrons into aromatic systems decreases their availability for hydrogen bonding with the

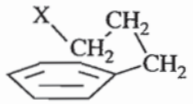
Table 2.6 Effect of Folding of Alkyl Chains on π[43]

X	π_X (aromatic)[a]	π_X (aliphatic)[b]	$\Delta\pi_X$
OH	−1.80	−1.16	0.64
F	−0.73	−0.17	0.56
Cl	−0.13	0.39	0.52
Br	0.04	0.60	0.56
I	0.22	1.00	0.78
COOH	−1.26	−0.67	0.59
CO_2CH_3	−0.91	−0.27	0.64
$COCH_3$	−1.26	−0.71	0.55
NH_2	−1.85	−1.19	0.66
CN	−1.47	−0.84	0.63
OCH_3	−0.98	−0.47	0.51
$CONH_2$	−2.28	−1.71	0.57
		Average	0.60 ± 0.05

[a] Log $P_{Ph(CH_2)_3X}$ − log $P_{Ph(CH_2)_3H}$.
[b] Log $P_{CH_3(CH_2)_3X}$ − log $P_{CH_3(CH_2)_3H}$.

aqueous phase and, therefore, increases the π. This is supported by the general trend that aromatic π_X values are greater than aliphatic π_X values, again emphasizing the constitutive nature of π and log P.

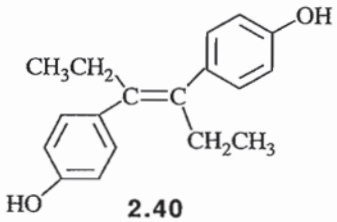
Steric effects are variable.[43] If a group sterically shields nonbonded electrons, then aqueous interactions will decrease, and the π value will increase. However, crowding of functional groups involved in hydrophobic interactions (see Chapter 3) will have the opposite effect. Conformational effects also can affect the π value.[43] The π_X values for $Ph(CH_2)_3X$ are consistently lower (more water soluble) than π_X values for $CH_3(CH_2)_3X$ (Table 2.6). This phenomenon is believed to be the result of folding of the side chain onto the phenyl ring (**2.39**), which means a smaller apolar surface for organic solvation. The folding may be caused by the interaction of the CH_2–X dipole with the phenyl π electrons and by intramolecular hydrophobic interactions.



2.39

Two examples follow to show the additivity of π constants in predicting log P values. A calculation of the log P for the anticancer drug diethylstilbestrol (**2.40**) is as follows:

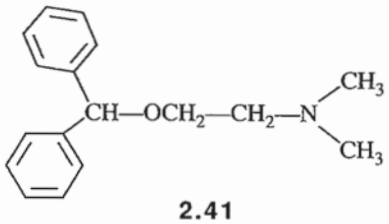
$$\begin{aligned} \text{Calc. log } P &= 2\pi_{CH_3} + 2\pi_{CH_2} + \pi_{CH=CH} + 2 \log P_{PhOH} - 0.40 \\ &= 2(0.50) + 2(0.50) + 0.69 + 2(1.46) - 0.40 \\ &= 5.21 \end{aligned} \qquad (2.11)$$



2.40

In Eq. (2.11), $\pi_{CH=CH} = \frac{1}{2}(\pi_{CH=CHCH=CH})$, which was shown in Table 2.5 to be $\frac{1}{2}(1.38)$; -0.40 is added into the equation to account for two branching points (each end of the alkene). The calculated log P value of 5.21 is quite remarkable considering that the experimental log P value is 5.07.

A calculation of log P for the antihistamine diphenhydramine (**2.41**) is shown in Eq. (2.12). In Eq. (2.12), 2.13 is log P for benzene, which is the same as π_{Ph}; 0.30 is $\pi_{CH_2}(0.50) - 0.20$ for branching; -0.73 was obtained by subtracting 1.50 ($2\pi_{CH_3} + \pi_{CH_2}$) from log $P_{CH_3CH_2OCH_2CH_3}$ (=0.77); and -0.95 is the value for π_{NMe_2} obtained from $Ph(CH_2)_3NMe_2$.[43] The experimental log P value is 3.27.



2.41

$$\begin{aligned} \text{Calc. log } P &= 2\pi_{Ph} + \pi_{CH} + \pi_{COH_2} + \pi_{CH_2} + \pi_{NMe_2} \\ &= 2(2.13) + 0.30 - 0.73 + 0.50 - 0.95 \\ &= 3.38 \end{aligned} \qquad (2.12)$$

The chore of calculating log P values for molecules has been lessened considerably by the computerization of the method.[45] A nonlinear regression model for the estimation of partition coefficients was developed by Bodor *et al.*[46] using the following molecular descriptors: molecular surface, volume, weight, and charge densities. It was shown to have excellent predictive power for the estimation of log P for complex molecules.

Although the log P values determined from 1-octanol/water partitioning are excellent models for *in vivo* lipophilicity, it has been found for a variety of aromatic compounds with log P values exceeding 5.5 (very lipophilic) or molar volumes greater than 230 cm^3/mol that there is a breakdown in the correlation of these values with those determined from partitioning between L-α-phosphatidylcholine dimyristoyl membrane vesicles and water.[47] Above a log P value of 5.5 the solvent solubility for these molecules is greater than their membrane solubility. As the compound increases in size more energy per unit volume is required to form a cavity in the structured membrane

phase. This is consistent with observations that branched molecules have lower log P values than their straight chain counterparts and that this effect is even greater in membranes than in organic solvents.

It should be noted that although log P values are most commonly determined with 1-octanol/water mixtures, this is not universal. For example, Seiler[48] introduced a new additive constitutive substituent constant for solvents other than 1-octanol. Therefore, when using log P values, it is important to be aware of the solvent used to obtain the log P data.

c. Steric Effects: The Taft Equation. Since interaction of a drug with a receptor involves the mutual approach of two molecules, another important parameter for QSAR is the *steric effect*. In much the same way that Hammett derived quantitative electronic effects (see Section II,E,2,a), Taft[49] defined the *steric parameter* E_s [Eq. (2.13)]. Taft used for the reference reaction the relative rates of the acid-catalyzed hydrolysis of α-substituted acetates (XCH_2CO_2Me). This parameter is normally standardized to the methyl group ($XCH_2 = CH_3$) so that $E_s(CH_3) = 0.0$; it is possible to standardize it to hydrogen by adding 1.24 to every methyl-based E_s value.[50] Hancock *et al.*[51] claimed that this model reaction was under the influence of hyperconjugative effects and, therefore, developed corrected E_s values for the hyperconjugation of α-hydrogen atoms [Eq. (2.14)], where E_s^c is the corrected E_s value and n is the number of α-hydrogen atoms.

$$E_s = \log k_{XCOMe} - \log k_{CH_3CO_2Me} = \log \frac{k_X}{k_o} \tag{2.13}$$

$$E_s^c = E_s + 0.306(n - 3) \tag{2.14}$$

Two other steric parameters worth mentioning are molar refractivity (MR) and the Verloop parameter. *Molar refractivity*[52] is defined by the Lorentz–Lorenz equation [Eq. (2.15)], where n is the index of refraction at the sodium D line, MW is the molecular weight, and d is the density of the compound. The greater the positive MR value of a substituent, the larger is its steric or bulk effect. This parameter also measures the electronic effect and, therefore, may reflect dipole–dipole interactions at the receptor site.

$$MR = \frac{n^2 - 1}{n^2 + 2} \frac{MW}{d} \tag{2.15}$$

The *Verloop steric parameters*[53] are used in a program called STERIMOL to calculate the steric substituent values from standard bond angles, van der Waals radii, bond lengths, and user-determined reasonable conformations. Five parameters are involved. One (L) is the length of the substituent along the axis of the bond between the substituent and the parent molecule. Four width parameters (B_1–B_4) are measured perpendicular to the bond axis. These

five parameters describe the positions, relative to the point of attachment and the bond axis, of five planes which closely surround the group. In contrast to E_s values which, because of the reaction on which they are based, cannot be determined for many substituents, the Verloop parameters are available for any substituent.

3. Methods Used to Correlate Physicochemical Parameters with Biological Activity

Now that we can obtain numerous physicochemical parameters (also called descriptors) for any substituent, how do we use these parameters to gain information regarding what compound to synthesize next in an attempt to optimize the lead compound? First, several (usually, many) compounds related to the lead are synthesized, and the biological activities are determined in some bioassay. These data, then, can be manipulated by a number of QSAR methods. The most popular is Hansch analysis.

a. Hansch Analysis: A Linear Multiple Regression Analysis. With the realization that there are (at least) two considerations for biological activity, namely, lipophilicity (required for the journey of the drug to the site of action) and electronic factors (required for drug interaction with the site of action), and that lipophilicity is a parabolic function, Hansch and Fujita[37] expanded Eq. (2.6) to that shown in either Eq. (2.16a) or (2.16b) known as the *Hansch equation*, where C is the molar concentration (or dose) that elicits a standard biological response (e.g., ED_{50}, the dose required for 50% of the maximal effect, IC_{50}, the concentration that gives 50% inhibition of an enzyme or antagonism of a receptor; or LD_{50}, the lethal dose for 50% of the animal population). The terms k, k', ρ, and k'' are the regression coefficients derived from statistical curve fitting, and π and σ are the lipophilicity and electronic substituent constants, respectively. The reciprocal of the concentration ($1/C$) reflects the fact that greater potency is associated with a lower dose, and the negative sign for the π^2 [or $(\log P)^2$] term reflects the expectation of an optimum lipophilicity, that is, the π_o or $\log P_o$.

$$\log 1/C = -k\pi^2 + k'\pi + \rho\sigma + k'' \qquad (2.16a)$$

$$\log 1/C = -k(\log P)^2 + k'(\log P) + \rho\sigma + k'' \qquad (2.16b)$$

Because of the importance of steric effects and other shape factors of molecules for receptor interactions, an E_s term and a variety of other shape, size, or topography terms (S) have been added to the Hansch equation [see Eq. (2.17)]. The way these parameters are used is by the application of the method of *linear multiple regression analysis*.[54] The best least squares fit of the dependent variable (the biological activity) to a linear combination of the independent variables (the descriptors) is determined. *Hansch analysis*, also

called the *extrathermodynamic method*, then, is a linear free energy approach to drug design in congeneric series in which equations are set up involving different combinations of the physicochemical parameters; the statistical methodology allows the best equation to be selected and the statistical significance of the correlation to be assessed. Once this equation has been established, it can be used to predict the activities of untested compounds. Problems associated with the use of multiple regression analysis in QSAR studies have been discussed by Deardon.[55]

$$\log 1/C = -a\pi^2 + b\pi + \rho\sigma + cE_s + dS + e \qquad (2.17)$$

Several assumptions must be made when the extrathermodynamic method is utilized: conformational changes in receptors can be ignored, metabolism does not interfere, linear free energy terms relevant to receptor affinity are additive, the potency–lipophilicity relationship is parabolic or linear, and correlation implies a causal relationship. According to Martin[56a] and Tute[56b] there is a balance of assets and liabilities to the extrathermodynamic method. The strengths are severalfold: (1) the use of descriptors (π, σ, E_s, MR, and so forth) permits data collected from simple organic chemical model systems to be utilized for the prediction of biological activity in complex systems; (2) the predictions are quantitative with statistical confidence limits; (3) the method is easy to use and is inexpensive; and (4) conclusions that are reached may have application beyond the substituents included in the particular analysis.

The weaknesses of this method are that (1) there must be parameter values available for the substituents in the data set; (2) a large number of compounds must be included in the analysis in order to have confidence in the derived equations; (3) expertise in statistics and computer use is essential; (4) small molecule interactions are imperfect models for biological systems; (5) in contrast to chemical reactions in which one knows the atoms that interact with the reagent, steric effects in biological systems may not be relevant, since it is often not certain which atoms in the drug interact with the receptor; (6) organic reactions used to determine the descriptors usually are studied under acidic or basic conditions when all analogs are fully protonated or deprotonated, whereas in biological systems the drug may be partially protonated; (7) since QSAR study is empirical, it is a retrospective technique that depends on the pharmacological activity of compounds belonging to the same structural type, and, therefore, new types of active compounds are not discovered (i.e., it is a lead optimization technique, not a lead discovery approach); and (8) like other empirical relationships, extrapolations frequently lead to false predictions.

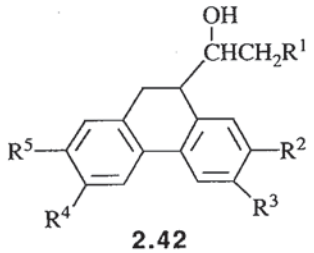
Despite the weaknesses of this approach it is used widely, and several successes in drug design attributable to Hansch analysis have been reported.[57] As pointed out in Section III,F,2 of Chapter 3, however, caution should be used when applying QSAR methods to racemic mixtures if only one enantio-

mer is active. Although Hansch analysis is the foremost method, there are other important statistical approaches that will be mentioned briefly.

b. Free and Wilson or de Novo Method. Not long after Hansch proposed the extrathermodynamic approach, Free and Wilson[58] reported a general mathematical method for assessing the occurrence of additive substituent effects, and for quantitatively estimating their magnitude. It is a method for the optimization of substituents within a given molecular framework that is based on the assumption that the introduction of a particular substituent at any one position in a molecule always changes the relative potency by the same amount, regardless of what other substituents are present in the molecule. A series of linear equations, constructed of the form shown in Eq. (2.18), where *BA* is the magnitude of the biological activity, X_i is the *i*th substituent with a value of 1 if present and 0 if not, a_i is the contribution of the *i*th substituent to the *BA*, and μ is the overall average activity of the parent skeleton, are solved by the method of least squares for the a_i and μ. All activity contributions at each position of substitution must sum to zero. The pros and cons of the Free–Wilson method have been discussed by Blankley.[59] Fujita and Ban[60] suggested two modifications of the Free–Wilson approach on the assumption that the effect on the activity of a certain substituent at a certain position in a compound is constant and additive. First, they suggested that the biological activity should be expressed as log A/A_o, where A and A_o represent the magnitude of the activity of the substituted and unsubstituted compounds, respectively, and that a_i is the log activity contribution of the *i*th substituent relative to H. This allows the derived substituent constants to be compared directly with other parameters related to free energy that are additive. Second, they suggested that μ become analogous to the theoretically predicted (calculated) activity of the parent compound of the series. Both of these modifications have been widely accepted.

$$BA = \sum a_i X_i + \mu \tag{2.18}$$

As an example of the Free–Wilson approach, consider the hypothetical compound **2.42**.[56a] If in one pair of analogs for which R^1, R^2, R^3, and R^4 are constant and R^5 is Cl or CH_3, the methyl compound is one-tenth as potent as the chloro analog, then the Free–Wilson method assumes that every R^5 methyl analog (where R^1–R^4 are varied) will be one-tenth as potent as the corresponding R^5 chloro analog. A requirement for this approach, then, is a series of compounds that have changes at more than one position. In addition, each type of substituent must occur more than once at each position in which it is found. The outcome is a table of the contribution to potency of each substituent at each position. If the free energy relationships of the extrathermodynamic method are linear or position specific, then Free–Wilson calculations will be successful.



2.42

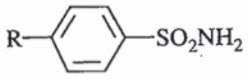
The *interaction model*[61] is a mathematical model similar to that of the Free–Wilson additive model with an additional term ($e_X e_Y$) to account for possible interactions between substituents X and Y.

c. Enhancement Factor. One of the earliest QSAR observations resulted from a retrospective analysis of a large number of synthetic corticosteroids.[62] Examination of the biological properties of steroids prepared by the introduction of halogen, hydroxyl, alkyl, or double bond modifications revealed that each substituent affects the activity of the molecule in a quantitative sense, and almost independently of other groups. The effect (whether positive or negative) of each substituent was assigned a numerical value termed the *enhancement factor*. Multiplication of the enhancement factor for each substituent by the biological activity of the unsubstituted compound gave the potency of the modified steroid.

d. Manual Stepwise Methods: Topliss Operational Schemes and Others. Since organic chemists are, by nature, more likely to be intuitive and less so mathematical, it was not long before Topliss[63] developed a nonmathematical, nonstatistical, and noncomputerized (hence, *manual*) guide to the use of the Hansch principles. This method is most useful when the synthesis of large numbers of compounds is difficult and when biological testing of compounds is readily available. It is an approach for the efficient optimization of the potency of a lead compound with minimization of the number of compounds needed to be synthesized. The only prerequisite for the technique is that the lead compound must contain an unfused benzene ring. However, according to literature surveys at the time that this method was published, 40% of all reported compounds[64] contain an unfused benzene ring and 50% of drug-oriented patents[65] are concerned with substituted benzenes. This approach relies heavily on π and σ values and to a much lesser degree E_s values. The methodology is outlined here; a more detailed discussion can be found in the Topliss paper.[63]

Consider that the lead compound is benzenesulfonamide (**2.43**, R = H) and its potency has been measured in whatever bioassay is being used. Since many systems are $+\pi$ dependent, that is, the potency increases with increasing π values, then a good choice for the first analog would be one with a

substituent having a $+\pi$ value. Since $\pi_{4\text{-Cl}} = 0.71$ and $\sigma_{4\text{-Cl}} = 0.23$ (remember, $\pi_H = \sigma_H = 0$), the 4-chloro analog (**2.43**, R = Cl) should be synthesized and tested. There are three possible outcomes of this effort, namely, the 4-chloro analog is more potent, equipotent, or less potent than the parent compound. If it is more potent, then it can be attributed to a $+\pi$ effect, a $+\sigma$ effect, or to both. In this case, the 3,4-dichloro analog ($\pi_{3,4\text{-Cl}_2} = 1.25$, $\sigma_{3,4\text{-Cl}_2} = 0.52$) could be synthesized next and tested. Again, the 3,4-dichloro analog could be more potent, equipotent, or less potent than the 4-chloro compound. If it is more potent, then determination of whether $+\pi$ or $+\sigma$ is more important could be made by selection next of the 4-SPh analog ($\pi_{SPh} = 2.32$, $\sigma_{SPh} = 0.18$) or the 3-trifluoromethyl-4-nitro analog ($\pi_{3\text{-CF}_3\text{-4-NO}_2} = 0.60$, $\sigma_{3\text{-CF}_3\text{-4-NO}_2} = 1.21$).



2.43

At this point a potency tree, termed a *Topliss decision tree*, could be constructed (Fig. 2.5), and additional analogs could be made. It must be stressed that this analysis was based solely on π and σ values, and other factors such as steric effects have been neglected.

If the 3,4-dichloro compound was less potent than the 4-chloro analog, it could be that the optimum values of π and σ were exceeded or that the 3-chloro group has an unfavorable steric effect. The latter hypothesis could be tested by the synthesis of the 4-trifluoromethyl analog ($\pi_{4\text{-CF}_3} = 0.88$, $\sigma_{4\text{-CF}_3} =$

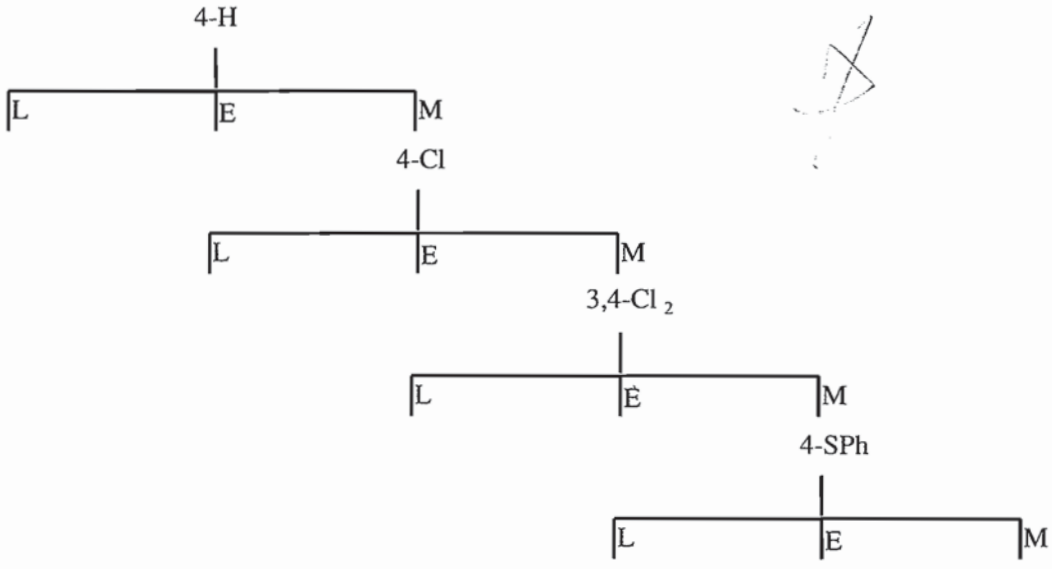


Figure 2.5. Topliss decision tree (M, more potent; E, equipotent; L, less potent).

0.54), which has no 3-substituent but has a high σ and intermediate π value. If this analog is more potent than the 4-chloro analog, the 4-nitro analog ($\pi_{4\text{-}NO_2} = -0.28$, $\sigma_{4\text{-}NO_2} = 0.78$) or the 4-ethyl analog ($\pi_{4\text{-Et}} = 1.02$, $\sigma_{4\text{-Et}} = -0.15$) could be synthesized in order to determine the importance of π and σ values, respectively.

What if the 4-chloro analog was equipotent with the parent compound? This could result from a favorable $+\pi$ effect counterbalanced by an unfavorable σ effect or vice versa. If the former is the case, then the 4-methyl analog ($\pi_{4\text{-Me}} = 0.56$, $\sigma_{4\text{-Me}} = -0.17$) should show enhanced potency. Further enhancement of potency by the 4-methyl analog would suggest that the synthesis of analogs with increasing π values and decreasing σ values would be propitious. If the 4-methyl analog is worse than the 4-chloro analog, perhaps the equipotency of the 4-chloro compound was the result of a favorable σ effect and an unfavorable π effect. The 4-nitro analog ($\pi_{4\text{-}NO_2} = -0.28$, $\sigma_{4\text{-}NO_2} = 0.78$) would, then, be a wise next choice.

If the 4-chloro analog was less potent than the lead, then there may be a steric problem at the 4 position, or increased potency may depend on $-\pi$ and $-\sigma$ values. The 3-chloro analog ($\pi_{3\text{-Cl}} = 0.71$, $\sigma_{3\text{-Cl}} = 0.37$) could be synthesized to determine if a steric effect is the problem. Note that the σ constant for the 3-Cl substituent is different from that for the 4-Cl one because these descriptors are constitutive. If there is no steric effect, then the 4-methoxy compound ($\pi_{4\text{-OMe}} = -0.04$, $\sigma_{4\text{-OMe}} = -0.27$) could be prepared to investigate the effect of adding a $-\sigma$ substituent. An increased potency of the 4-OMe substituent would suggest that other substituents with more negative π and/or σ constants be tried.

Topliss[63] extended the operational scheme for side-chain problems when the group is adjacent to a carbonyl, amino, or amide functionality, namely, —COR, —NHR, —CONHR, —NHCOR, where R is the variable substituent. This approach is applicable to a variety of situations other than direct substitution on the aromatic nucleus. In this case, the parent molecule is the one where R is CH_3, and π, σ, and E_s parameters are used.

Note that in the Topliss operational scheme, as for other methods in this section, the procedure is *stepwise*, that is, the next compound is determined on the basis of the results obtained with the previous one. Three other manual, stepwise methods are mentioned only briefly: Craig plots,[66] Fibonacci search method,[67] and sequential simplex strategy.[68] The Topliss decision tree approach evolved from the work of Craig,[66] who pointed out the utility of a simple graphical plot of π versus σ (or any two parameters) to guide the choice of a substituent (Fig. 2.6). Once the Hansch equation has been expressed for an initial set of compounds, the sign and magnitude of the π and σ regression coefficients determine the particular quadrant of the Craig plot that is to be used to direct further synthesis. Thus, if both the π and σ terms have

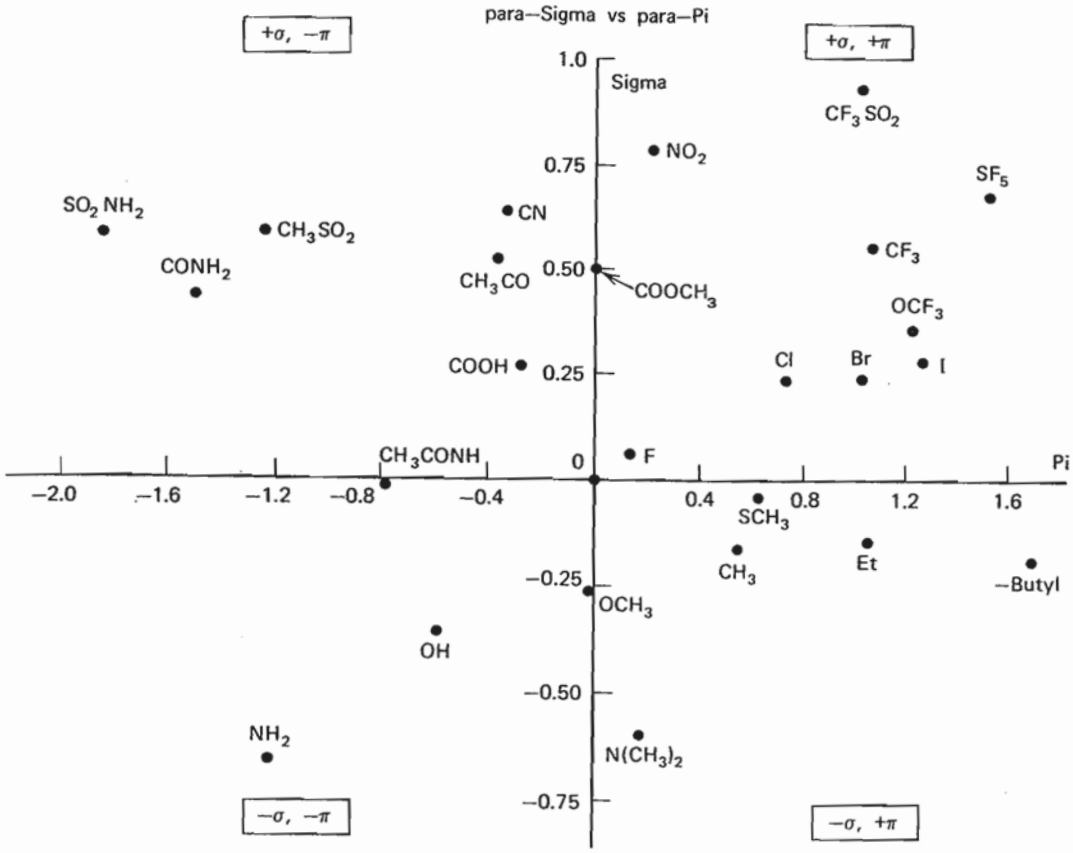


Figure 2.6. Craig plot of σ constants versus π values for aromatic substituents.[66] [Reprinted by permission of John Wiley & Sons, Inc. from Craig, P. N. (1980). *In* "Burger's Medicinal Chemistry," (M. E. Wolff, ed.), 4th ed., Part I, p. 343. Wiley, New York. Copyright © 1980 John Wiley & Sons, Inc.]

positive coefficients, then substituents in the upper right-hand quadrant of the plot (Fig. 2.6) should be selected for future analogs.

The Fibonacci search technique[67] is a manual method to discover the optimum of some parabolic function, such as potency versus log P, in a minimum number of steps. Sequential simplex strategy[68] is another stepwise technique suggested when potency depends on two physicochemical parameters such as π and σ.

e. Batch Selection Methods: Batchwise Topliss Operational Scheme, Cluster Analysis, and Others. The inherent problem with the Topliss operational scheme described in Section II,E,3,d is its stepwise nature. Provided that pharmacological results can be obtained quickly, this is probably not much of a problem; however, biological evaluation is often slow. Topliss[69] proposed an alternative scheme that uses *batchwise* analysis of small groups of com-

Table 2.7 Potency Order for Various Parameter Dependencies [With permission from Topliss, J. G. (1977). *J. Med. Chem* **20,** 463. Copyright © 1977 American Chemical Society.]

	Parameters									
Substituent	π	$2\pi - \pi^2$	σ	$-\sigma$	$\pi + \sigma$	$2\pi - \sigma$	$\pi - \sigma$	$\pi - 2\sigma$	$\pi - 3\sigma$	E_4[a]
3,4-Cl_2	1	1–2	1	5	1	1	1–2	3–4	5	2–5
4-Cl	2	1–2	2	4	2	2–3	3	3–4	3–4	2–5
4-CH_3	3	3	4	2	3	2–3	1–2	1	1	2–5
4-OCH_3	4–5	4–5	5	1	5	4	4	2	2	2–5
H	4–5	4–5	3	3	4	5	5	5	3–4	1

[a] Unfavorable steric effect from 4-substitution.

pounds. Substituents were grouped by Topliss[69] according to π, σ, π^2, and a variety of $x\pi$ and $y\sigma$ weighted combinations. The approach starts with the synthesis of five derivatives, the unsubstituted (4-H), 4-chloro, 3,4-dichloro, 4-methyl, and 4-methoxy compounds. After these five analogs have been tested in the bioassay, they are ranked in order of decreasing potency. The potency order determined for these analogs is then compared with the rankings in Table 2.7 to determine which parameter or combination of parameters is most dominant. If, for example, the potency order is 4-OCH_3 > 4-CH_3 > H > 4-Cl > 3,4-Cl_2, then $-\sigma$ is the dominant parameter. Once the parameter dependency is determined, Table 2.8 is consulted in order to discover what substituents should be investigated next. In the above example, 4-$N(C_2H_5)_2$, 4-$N(CH_3)_2$, 4-NH_2, 4-NHC_4H_9, 4-OH, 4-$OCH(CH_3)_2$, 3-CH_3, and 4-OCH_3 would be suitable choices. The major weakness of this approach is that it is

Table 2.8 New Substituent Selections [With permission from Topliss, J. G. (1977). *J. Med. Chem.* **20,** 463. Copyright © 1977 American Chemical Society.]

Probable operative parameters	New substituent selection
π, $\pi + \sigma$, σ	3-CF_3, 4-Cl; 3-CF_3, 4-NO_2; 4-CF_3, 2,4-Cl_2; 4-*c*-C_5H_9; 4-*c*-C_6H_{11}
π, $2\pi - \sigma$, $\pi - \sigma$	4-$CH(CH_3)_2$; 4-$C(CH_3)_3$; 3,4-$(CH_3)_2$; 4-$O(CH_2)_3CH_3$; 4-OCH_2Ph; 4-$N(C_2H_5)_2$
$\pi - 2\sigma$, $\pi - 3\sigma$, $-\sigma$	4-$N(C_2H_5)_2$; 4-$N(CH_3)_2$; 4-NH_2; 4-NHC_4H_9; 4-OH; 4-$OCH(CH_3)_2$; 3-CH_3, 4-OCH_3
$2\pi - \pi^2$	4-Br; 3-CF_3; 3,4-$(CH_3)_2$; 4-C_2H_5; 4-$O(CH_2)_2CH_3$; 3-CH_3, 4-Cl; 3-Cl; 3-CH_3; 3-OCH_3; 3-$N(CH_3)_2$; 3-CF_3; 3,5-Cl_2
Ortho effect	2-Cl; 2-CH_3; 2-OCH_3; 2-F
Other	4-F; 4-$NHCOCH_3$; 4-$NHSO_2CH_3$; 4-NO_2; 4-$COCH_3$; 4-SO_2CH_3; 4-$CONH_2$; 4-SO_2NH_2

Table 2.9 Typical Members of Clusters Based on α, π^2, *F*, *R*, *MR*, and *MW* [Reprinted with permission from Martin, Y. C. (1979). *In* "Drug Design" (E. J. Ariens, ed.), Vol. VIII, p. 5. Academic Press, New York. Copyright © 1979 Academic Press, Inc.]

Cluster number[a]	Typical members
1	Me, H, 3,4-(OCH_2O), CH_2CH_2COOH, $CH{=}CH_2$, Et, CH_2OH
2	$CH{=}CHCOOH$
3a	CN, NO_2, CHO, COOH, COMe
3b	$C{\equiv}CH$, CH_2Cl, Cl, NNN, SH, SMe, $CH{=}NOH$, CH_2CN, OCOMe, SCOMe, COOMe, SCN
4a	$CONH_2$, CONHMe, SO_2NH_2, SO_2Me, SOMe
4b	NHCHO, NHCOMe, $NHCONH_2$, $NHCSNH_2$, $NHSO_2Me$
5	F, OMe, NH_2, $NHNH_2$, OH, NHMe, NHEt, NMe_2
6	Br, OCF_3, CF_3, NCS, I, SF_5, SO_2F
7	CH_2Br, SeMe, $NHCO_2Et$, SO_2Ph, OSO_2Me
8	NHCOPh, $NHSO_2Ph$, OSO_2Ph, COPh, $N{=}NPh$, OCOPh, PO_2Ph
9	3,4-$(CH_2)_3$, 3,4-$(CH_2)_4$, Pr, *i*-Pr, 3,4-$(CH)_4$, NHBu, Ph, CH_2Ph, *t*-Bu, OPh
10	Ferrocenyl, adamantyl

[a] Clusters 3 and 4 contain many of the common substituents used in medicinal chemistry; hence, these clusters are further subdivided according to their cluster membership when 20 clusters have been made.

difficult to extend the method to additional parameters unless computers are utilized.

A computer-based batch selection method, known as *cluster analysis*, was introduced by Hansch *et al.*[70] Substituents were grouped into clusters with similar properties according to their σ, π, π^2, E_s, *F* (field constant), *R* (resonance constant), *MR* (molar refractivity), and MW (molecular weight) values. Some of the clusters are shown in Table 2.9.[71] One member of each cluster would be selected for substitution into the lead compound, and the compounds would be synthesized and tested. If a substituent showed dominant potency, then other substituents from that cluster would be selected for further investigation. The important advantage of the batch selection methods is that the initial batch of analogs prepared is derived from the widest range of parameters possible so that the dominant physicochemical property can be revealed early in the lead modification process.

There are other selection methods and statistical analysis techniques, but they lie outside the scope of this book (see General References, this chapter).

4. Computer-Based Methods of QSAR Related to Receptor Binding

There are a variety of computer-based methods that have been used to correlate molecular structure with receptor binding, and, therefore, activity. Some

are mentioned here. Crippen and co-workers[72a–c] devised a linear free energy model, termed the *distance geometry approach*, for calculating QSAR from receptor binding data. The distances between various atoms in the molecule, compiled into a table called the *distance matrix*, define the conformation of the molecule. Rotations about single bonds change the molecular conformation and, therefore, these distances; consequently, an upper and lower distance limit is set on each distance. Experimentally determined free energies of binding of a series of compounds to the receptor are used with the distance matrix of each molecule in a computerized method to deduce possible binding sites in terms of geometry and chemical character of the site. Although this approach requires more computational effort and adjustable parameters than Hansch analysis, it is suggested[72b] to give good results on more difficult data sets.

The distance geometry approach was extended by Sheridan *et al.*[73] to treat two or more molecules as a single ensemble. The *ensemble approach to distance geometry* can be used to find a common pharmacophore for a receptor with unknown structure from a small set of biologically active molecules.

Hopfinger[74] has developed a set of computational procedures termed *molecular shape analysis* for the determination of the active conformations and, thereby, molecular shapes during receptor binding. Common pairwise overlap steric volumes calculated from low-energy conformations of molecules are used to obtain three-dimensional molecular shape descriptors which can be treated quantitatively and used with other physicochemical parameter descriptors.

Two other descriptors for substructure representation, the *atom pair*[75] and the *topological torsion*,[76] have been described by Venkataraghavan and coworkers. These descriptors characterize molecules in fundamental ways that are useful for the selection of potentially active compounds from hundreds of thousands of structures in a database. The atom pair method can select compounds from diverse structural classes that have atoms within the entire molecule similar to those of a particular active structure. The topological torsion descriptor is complementary to the atom pair descriptor, and it focuses on a local environment of a molecule for comparison with active structures.

F. Molecular Graphics-Based Drug Design

Quantitative structure–activity relationship studies have relied heavily on the use of computers from the beginning for statistical calculations involving multiparameter equations. It was soon realized that drug design could be aided significantly if structures of receptors and drugs could be displayed on a terminal and molecular processes could be visualized. *Molecular graphics* is the use of computer graphics to represent and manipulate molecular struc-

tures. The origins of molecular graphics has been traced by Hassall[77] to the project MAC (Multiple Access Computer),[78] which produced computer graphics models of macromolecules for the first time. The potential to apply this technology to protein crystallography was quickly realized, and by the early 1970s electron density data from X-ray diffraction studies could be presented and manipulated in stick or space-filling multicolored representations on a computer terminal.[79]

Medicinal chemists saw the potential of this approach in drug design as well. Stick (Dreiding) and space-filling (CPK) molecular models have been used extensively by organic chemists for years, but these hand-held models have major disadvantages.[80] Space-filling models often obscure the structure of the molecule, and wire or plastic models can give false impressions of molecular flexibility and often tend to change into unfavorable conformations at inopportune moments. A three-dimensional computer graphics representation that can be manipulated in three dimensions allows the operator to visualize the interactions of small molecules with biologically important macromolecules. Superimposition of structures, which is cumbersome at best with manual models, can be performed easily by molecular graphics. Also, some systems have the capability to synthesize graphically new structures by the assemblage of appropriate molecular fragments from a fragment file.

There are numerous molecular graphics systems available,[80,81a,b] but the typical system, utilized by every major pharmaceutical company in the United States, Western Europe, and Japan, consists of a mainframe or supermini computer linked to a high-resolution graphics terminal with local intelligence. The graphics terminal may be equipped with a variety of peripheral devices such as graphic tablets, light pens, function keys, and dials to effect the molecular display and three-dimensional manipulations. The mainframe or minicomputer executes all of the molecular calculations, such as calculations of bond lengths, bond angles, and quantum chemical or force field calculations.

Once the computer graphics system is set up, there are a variety of approaches that can be taken to utilize it for drug design. The basic premise in the utilization of molecular graphics is that the better the complementary fit of the drug to the receptor, the more potent the drug will be. This is the *lock-and-key hypothesis* of Fischer[82] in which the receptor is the lock into which the key (i.e., the drug) fits. In order to apply this concept it would appear that the structure of the receptor would have to be known, then different drug analogs could be docked into the receptor. *Docking* is a molecular graphics term for the computer-assisted movement of a terminal-displayed molecule into its receptor. However, as indicated in Chapter 3, the structures of very few receptors, except for enzymes, are known. This methodology, then, would appear to be quite limited; however, the ability to accomplish this would be a simple and useful drug design approach. Kuntz *et al.*[83] reported an algorithm

designed to fit small molecules into their macromolecular receptors. This shape-matching method, which is restricted to rigid *ligands* (receptor-bound molecules) and receptors, was modified[84] for flexible ligands where a ligand is approximated as a small set of rigid fragments. The drawbacks of this approach are the assumptions that binding is determined primarily by shape complementarity and that only small changes in the shape of the receptor occur upon ligand binding. An important advantage, though, is that this method is not limited to docking of known ligands. A library of molecular shapes can be scanned to determine which shapes best fit a particular receptor binding site.

Since the energetics, as well as the shape complementarity, of a drug–receptor complex are vital to its stability, Goodford[85] described a method that simultaneously displays the energy contour surfaces and the macromolecular structure on the computer graphics system. This allows both the energy and shape to be considered together when considering the design of molecules that have the optimal fit into the receptor.

There are relatively few receptors with known X-ray structures. Consequently, it would appear that the approach of *receptor fitting* is of little importance to drug design. However, because of the great advances in computer technology and software development,[81b] it is possible to use molecular graphics to obtain information about the ligand binding site of an unknown receptor. One approach is to deduce the topography of an unknown receptor site from related known receptor structures.[86] Another approach is to use molecular graphics visualization of an electron density map for a known drug–receptor complex obtained by X-ray crystallography. This may reveal empty pockets in the complex that could be filled by appropriate modification, an approach taken by Blaney *et al.*[87] in the design of new thyroid hormone analogs.

A third important molecular graphics technique useful for identification of the pharmacophore geometry is called *receptor mapping*. This method utilizes data from known ligand binding studies to an unknown receptor. With this technique, which also is founded on the premise that receptor topography is complementary to that of drugs, the structure of the lock is deduced from the shape of the keys that fit it. A variety of receptor mapping techniques have been described. An approach termed *steric mapping*[88] uses molecular graphics to combine the volumes of compounds known to bind to the desired receptor. This composite volume generates an *enzyme-excluded volume map* which defines that region of the binding site available for binding by drug analogs and, therefore, not occupied by the receptor itself. The same procedure is then carried out for similar molecules that are inactive. The composite volume is inspected for regions of volume overlap common to all of the inactive analogs. These are the *enzyme-essential regions*, sites required by the receptor itself and unavailable for occupancy by ligands. Any other molecule that overlaps with these regions should be inactive. Drug design, then, would involve the synthesis of compounds with the appropriate pharma-

cophore that filled the enzyme-excluded regions and that avoided the enzyme-essential regions.

In the modified method for docking flexible ligands into a receptor[84] described above, an X-ray structure of the receptor is not necessary to characterize the shape of the receptor binding site. Rather, the receptor binding site can be deduced from the shapes of active ligands.

Although molecular graphics approaches are widely used, it is not clear if any drugs have yet been designed *de novo* by this method, although leads have been discovered and lead optimization has been assisted by these techniques.[77,89] Several problems with this approach may contribute to its less than optimal effectiveness. The major problem derives from the fact that pharmacokinetics are ignored by this method. Prior to the drug candidate interacting with a receptor, it must be properly absorbed, it must reach the receptor without metabolic or chemical degradation (unless it is a prodrug; see Chapter 8), excretion must be appropriate, and the drug candidate and metabolites must not be toxic nor lead to undesirable side effects. Another problem with molecular graphics approaches is that energy minimization programs[90] are used to determine the lowest energy conformers of molecules, and the calculations are carried out for ground state molecules; however, as will be discussed in Chapters 3 and 5, the conformation of a molecule during the receptor binding process is not necessarily the one having the lowest energy, and it can be quite different from the ground state conformation. Also, these calculations generally are performed on molecules in the absence of solvent effects. Finally, programs written for drug–receptor interactions assume that the receptor is reasonably rigid during binding, which may not be the case.

G. Epilogue

On the basis of what was discussed in this chapter, it appears that even if one uncovers a lead, it may be a fairly random process to optimize its potency. In fact, less than 1 in 10,000 compounds synthesized in drug companies makes it to the drug market, and, in so doing, it takes about 10 years of research at a cost of $200–250 million. However, more rational approaches to lead discovery and lead optimization, based on chemical and biochemical principles, can be used. Some of these approaches are discussed in Chapters 3, 5, and 6.

References

1. Fleming, A. 1929. *Br. J. Exp. Pathol.* **10,** 226.
2. Hare, R. 1970. "The Birth of Penicillin." Allen & Unwin, London.
3. Beveridge, W. I. B. 1981. "Seeds of Discovery." Norton, New York.

4. Abraham, E. P., Chain, E., Fletcher, C. M., Gardner, A. D., Heatley, N. G., Jennings, M. A., and Florey, H. W. 1941. *Lancet 2,* 177.
5. Florey, H. W., Chain, E., Heatley, N. G., Jennings, M. A., Sanders, A. G., Abraham, E. P., and Florey, M. E. 1949. "Antibiotics," Vol. 2. Oxford Univ. Press, London.
6. Moyer, A. J., and Coghill, R. D. 1946. *J. Bacteriol.* **51,** 79.
7. Sternbach, L. H. 1979. *J. Med. Chem.* **22,** 1.
8. Shen, T. Y. 1976. *In* "Clinoril in the Treatment of Rheumatic Disorders" (Huskisson, E. C., and Franchimont, P., eds.), Raven, New York.
9. Gay, L. N., and Carliner, P. E. 1949. *Science* **109,** 359.
10. Strickland, B. A., Jr., and Hahn, G. L. 1949. *Science* **109,** 359.
11. Shen, T. Y., and Winter, C. A. 1977. *Adv. Drug Res.* **12,** 89.
12. Grewe, R. 1946. *Naturwissenschaften* **33,** 333.
13. Schnider, O., and Grüssner, A. 1949. *Helv. Chim. Acta* **32,** 821.
14. May, E. L., and Murphy, J. G. 1955. *J. Org. Chem.* **20,** 257.
15. Schaumann, O. 1940. *Naunyn-Schmiedeberg's Arch. Pharmacol. Exp. Pathol.* **196,** 109.
16. Bentley, K. W., and Hardy, D. G. 1967. *J. Am. Chem. Soc.* **89,** 3267; Bentley, K. W., Hardy, D. G., and Meek, B. 1967. *J. Am. Chem. Soc.* **89,** 3273.
17. Crum-Brown, A., and Fraser, T. R. 1868–1869. *Trans. R. Soc. Edinburgh* **25,** 151 and 693; Crum-Brown, A., and Fraser, T. R. 1869. *Proc. R. Soc. Edinburgh* **6,** 556; Crum-Brown, A., and Fraser, T. R. 1872. *Proc. R. Soc. Edinburgh* **7,** 663.
18. Richardson, B. W. 1869. *Med. Times Gaz.* **18,** 703.
19. Northey, E. H. 1948. "The Sulfonamides and Allied Compounds." American Chemical Society Monograph Series, Reinhold, New York.
20. Loubatieres, A. 1969. *In* "Oral Hypoglycemic Agents" (Campbell, G. D., ed.). Academic Press, New York.
21. Sprague, J. M. 1968. *In* "Topics in Medicinal Chemistry" (Robinowitz, J. L., and Myerson, R. M., eds.), Vol. 2. Wiley, New York.
22a. Dohme, A. R. L., Cox, E. H., and Miller, E. 1926. *J. Am. Chem. Soc.* **48,** 1688.
22b. Funcke, A. B. H., Ernsting, M. J. E., Rekker, R. F., and Nauta, W. T. 1953. *Arzneim.-Forsch.* **3,** 503.
23. Thornber, C. W. 1979. *Chem. Soc. Rev.* **8,** 563.
24. Burger, A. 1970. *In* "Medicinal Chemistry" (Burger, A., ed.), 3rd Ed., Wiley, New York.
25. Korolkovas, A. 1970. "Essentials of Molecular Pharmacology: Background for Drug Design," p. 54. Wiley, New York.
26. Lipinski, C. A. 1986. *Annu. Rep. Med. Chem.* **21,** 283.
27. Grimm, H. G. 1925. *Z. Elektrochem.* **31,** 474; Grimm, H. G. 1928. *Z. Elektrochem.* **34,** 430.
28. Erlenmeyer, H. 1948. *Bull. Soc. Chim. Biol.* **30,** 792.
29. Hansch, C., Maloney, P. P., Fujita, T., and Muir, R. M. 1962. *Nature (London)* **194,** 178.
30. Richet, M. C. 1893. *C. R. Seances Soc. Biol. Ses Fil.* **45,** 775.
31. Overton, E. 1897. *Z. Phys. Chem.* **22,** 189.
32. Meyer, H. 1899. *Arch. Exp. Pathol. Pharmacol.* **42,** 109.
33. Ferguson, J. 1939. *Proc. R. Soc. London, Ser. B* **127,** 387.
34. Hansch, C., Muir, R. M., and Metzenberg, R. L., Jr. 1951. *Plant Physiol.* **26,** 812.
35. Hammett, L. P. 1940. "Physical Organic Chemistry." McGraw-Hill, New York.
36. Roberts, J. D., and Caserio, M. C. 1977. "Basic Principles of Organic Chemistry," 2nd Ed., p. 1331. Benjamin, Menlo Park, California.
37. Hansch, C., and Fujita, T. 1964. *J. Am. Chem. Soc.* **86,** 1616.
38. Singer, S. J., and Nicolson, G. L. 1972. *Science* **175,** 720.
39. Fujita, T., Iwasa, J., and Hansch, C. 1964. *J. Am. Chem. Soc.* **86,** 5175.
40. Collander, R. 1954. *Physiol. Plant* **7,** 420; Collander, R. 1951. *Acta Chem. Scand.* **5,** 774.
41. Hansch, C., Steward, A. R., Anderson, S. M., and Bentley, D. 1968. *J. Med. Chem.* **11,** 1.

42. Hansch, C., and Leo, A. 1979. "Substituent Constants for Correlation Analysis in Chemistry and Biology." Wiley, New York.
43. Leo, A., Hansch, C., and Elkins, D. 1971. *Chem. Rev.* **71,** 525.
44. Iwasa, J., Fujita, T., and Hansch, C. 1965. *J. Med. Chem.* **8,** 150.
45. Chou, J. T., and Jurs, P. C. 1979. *J. Chem. Inf. Comput. Sci.* **19,** 171; Pomona College Medicinal Chemistry Project; see Hansch, C., Björkroth, J. P., and Leo, A. 1987. *J. Pharm. Sci.* **76,** 663.
46. Bodor, N., Gabanyi, Z., and Wong, C.-K. 1989. *J. Am. Chem. Soc.* **111,** 3783 and 8062.
47. Gobas, F. A. P. C., Lahittete, J. M., Garofalo, G., Shiu, W. Y., and Mackay, D. 1988. *J. Pharm. Sci.* **77,** 265.
48. Seiler, P. 1974. *Eur. J. Med. Chem.* **9,** 473.
49. Taft, R. W. 1956. *In* "Steric Effects in Organic Chemistry" (Neuman, M. S., ed.), p. 556. Wiley, New York.
50. Unger, S. H., and Hansch, C. 1976. *Prog. Phys. Org. Chem.* **12,** 91.
51. Hancock, C. K., Meyers, E. A., and Yager, B. J. 1961. *J. Am. Chem. Soc.* **83,** 4211.
52. Hansch, C., Leo, A., Unger, S. H., Kim, K. H., Nikaitani, D., and Lien, E. J. 1973. *J. Med. Chem.* **16,** 1207.
53. Verloop, A., Hoogenstraaten, W., and Tipker, J. 1976. *In* "Drug Design" (Ariens, E. J., ed.), Vol. 7, p. 165. Academic Press, New York.
54. Daniel, C., and Wood, F. S. 1971. "Fitting Equations to Data." Wiley, New York; Draper, N. R., and Smith, H. 1966. "Applied Regression Analysis." Wiley, New York; Snedecor, G. W., and Cochran, W. G. 1967. "Statistical Methods." Iowa State Univ. Press, Ames, Iowa.
55. Deardon, J. C. 1987. *In* "Trends in Medicinal Chemistry" (Mutschler, E., and Winterfeldt, E., eds.), p. 109. VCH, Weinheim.
56a. Martin, Y. C. 1978. "Quantitative Drug Design: A Critical Introduction," Chap. 2. Dekker, New York.
56b. Tute, M. S. 1980. *In* "Physical Chemical Properties of Drugs" (Yalkowsky, S. H., Sinkula, A. A., and Valvani, S. C., eds.), p. 141. Dekker, New York.
57. Unger, S. H. 1984. *In* "QSAR in Design of Bioactive Compounds" (Kuchar, M., ed.), p. 1. Prous, Barcelona; Hopfinger, A. J. 1985. *J. Med. Chem.* **28,** 1133; Fujita, T. 1984. *In* "Drug Design: Fact or Fantasy?" (Jolles, G., and Wooldridge, K. R. H., eds.), Chap. 2. Academic Press, London.
58. Free, S. M., Jr., and Wilson, J. W. 1964. *J. Med. Chem.* **7,** 395.
59. Blankley, C. J. 1983. *In* "Quantitative Structure–Activity Relationships of Drugs" (Topliss, J. G., ed.), Chap. 1. Academic Press, New York.
60. Fujita, T., and Ban, T. 1971. *J. Med. Chem.* **14,** 148.
61. Bocek, K., Kopecký, J., Krivucová, M., and Vlachová, D. 1964. *Experientia* **20,** 667; Kopecký, J., Bocek, K., and Vlachová, D. 1965. *Nature* (*London*) **207,** 981.
62. Fried, J., and Borman, A. 1958. *Vitam. Horm.* (*N.Y.*) **16,** 303.
63. Topliss, J. G. 1972. *J. Med. Chem.* **15,** 1006.
64. Granito, C. E., Becker, G. T., Roberts, S., Wiswesser, W. J., and Windlinz, K. J. 1971. *J. Chem. Soc.* **11,** 106.
65. Goodford, P. J. 1973. *Adv. Pharmacol. Chemother.* **11,** 51.
66. Craig, P. N. 1971. *J. Med. Chem.* **14,** 680; Craig, P. N. 1980. *In* "Burger's Medicinal Chemistry" (Wolff, M. E., ed.), 4th Ed., Part 1, Chap. 8. Wiley, New York.
67. Bustard, T. M. 1974. *J. Med. Chem.* **17,** 777; Santora, N. J., and Auyang, K. 1975. *J. Med. Chem.* **18,** 959; Deming, S. N. 1976. *J. Med. Chem.* **19,** 977.
68. Darvas, F. 1974. *J. Med. Chem.* **17,** 799.
69. Topliss, J. G. 1977. *J. Med. Chem.* **20,** 463.
70. Hansch, C., Unger, S. H., and Forsythe, A. B. 1973. *J. Med. Chem.* **16,** 1217.

71. Martin, Y. C. 1979. *In* "Drug Design" (Ariens, E. J., ed.), Vol. 8, p. 5. Academic Press, New York.
72a. Crippen, G. M. 1979. *J. Med. Chem.* **22,** 988.
72b. Ghose, A. K., and Crippen, G. M. 1982. *J. Med. Chem.* **25,** 892.
72c. Crippen, G. M. 1981. "Distance Geometry and Conformational Calculations." Research Studies Press, New York.
73. Sheridan, R. P., Nilakantan, R., Dixon, J. S., and Venkataraghavan, R. 1986. *J. Med. Chem.* **29,** 899; Sheridan, R. P., and Venkataraghaven, R. 1987. *Acc. Chem. Res.* **20,** 322.
74. Hopfinger, A. J. 1980. *J. Am. Chem. Soc.* **102,** 7196; Hopfinger, A. J. 1981. *J. Med. Chem.* **24,** 818; Hopfinger, A. J. 1983. *J. Med. Chem.* **26,** 990.
75. Carhart, R. E., Smith, D. H., and Venkataraghavan, R. 1985. *J. Chem. Inf. Comput. Sci.* **25,** 64.
76. Nilakantan, R., Bauman, N., Dixon, J. S., and Venkataraghavan, R. 1987. *J. Chem. Inf. Comput. Sci.* **27,** 82.
77. Hassall, C. H. 1985. *Chem. Brit.* **21,** 39.
78. Levinthal, C. 1966. *Sci. Am.* **214(6),** 42.
79. Barry, C. D., and North, A. C. T. 1971. *Cold Spring Harbor Symp. Quant. Biol.* **36,** 577; Barry, C. D. 1971. *Nature* (*London*) **232,** 236.
80. Tollenaere, J. P., and Janssen, P. A. 1988. *J. Med. Res. Rev.* **8,** 1.
81a. Gund, P., Halgren, T. A., and Smith, G. M. 1987. *Annu. Rep. Med. Chem.* **22,** 269.
81b. Gund, T., and Gund, P. 1987. *In* "Molecular Structure and Energetics" (Liebman, J. F., and Greenberg, A., eds.), Vol. 4, Chap. 10. VCH, Weinheim.
82. Fischer, E. 1894. *Ber. Dtsch. Chem. Ges.* **27,** 2985.
83. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. 1982. *J. Mol. Biol.* **161,** 269.
84. DesJarlais, R. L., Sheridan, R. P., Dixon, J. S., Kuntz, I. D., and Venkatarghavan, R. 1986. *J. Med. Chem.* **29,** 2149; DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D., and Venkataraghavan, R. 1988. *J. Med. Chem.* **31,** 722.
85. Goodford, P. J. 1985. *J. Med. Chem.* **28,** 849.
86. Carlson, G. M., MacDonald, R. J., and Meyer, E. F., Jr. 1986. *J. Theor. Biol.* **119,** 107.
87. Blaney, J. M., Jorgensen, E. C., Connolly, M. L., Ferrin, T. E., Langridge, R., Oatley, S. J., Burridge, J. M., and Blake, C. C. F. 1982. *J. Med. Chem.* **25,** 785.
88. Sufrin, J. R., Dunn, D. A., and Marshall, G. R. 1981. *Mol. Pharmacol.* **19,** 307; Humblet, C., and Marshall, G. R. 1980. *Annu. Rep. Med. Chem.* **15,** 267; Marshall, G. R. 1985. *Ann. N.Y. Acad. Sci.* **439,** 162.
89. Hopfinger, A. J. 1985. *J. Med. Chem.* **28,** 1133.
90. Burkert, U., and Allinger, N. L. 1982. "Molecular Mechanics." American Chemical Society, Washington, D.C.; Boyd, D. B., and Lipkowitz, K. B. 1982. *J. Chem. Educ.* **59,** 269.

General References

Membranes and Membrane Biochemistry

Gennis, R. B. 1989. "Biomembranes: Molecular Structure and Function." Springer-Verlag, New York.
Jain, M. K., Wagner, R. C. 1980. "Introduction to Biological Membranes." Wiley, New York.
Vance, D. E., and Vance, J. E., eds. 1985. "Biochemistry of Lipids and Membranes." Benjamin/Cummings, Menlo Park, California.
Yeagle, P. 1987. "The Membranes of Cells." Academic Press, Orlando, Florida.