Biological Correlations—The Hansch Approach

A symposium sponsored by the Division of Pesticide Chemistry at the 161st Meeting of the American Chemical Society, Los Angeles, Calif., March 31–April 1, 1971.

Wade Van Valkenburg,

Symposium Chairman

ADVANCES IN CHEMISTRY SERIES 114

AMERICAN CHEMICAL SOCIETY
WASHINGTON, D. C. 1972

ADCSAj 114 1-304 (1972)

Library of Congress Catalog Card 72-95972

ISBN 8412-0156-0

PRINTED IN THE UNITED STATES OF AMERICA

Advances in Chemistry Series

Robert F. Gould, *Editor*

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.fw001

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.fw001

FOREWORD

ADVANCES IN CHEMISTRY SERIES was founded in 1949 by the American Chemical Society as an outlet for symposia and collections of data in special areas of topical interest that could not be accommodated in the Society's journals. It provides a medium for symposia that would otherwise be fragmented, their papers distributed among several journals or not published at all. Papers are refereed critically according to ACS editorial standards and receive the careful attention and processing characteristic of ACS publications. Papers published in ADVANCES IN CHEMISTRY SERIES are original contributions not published elsewhere in whole or major part and include reports of research as well as reviews since symposia may embrace both types of presentation.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

PREFACE

In the last decade the costs of introducing a new biological chemical (drug or pesticide) to market have nearly doubled. Current introductory costs are on the order of seven to 10 million dollars. As a result, fewer new drugs and pesticides are being marketed each year. Concurrently, many large industrial concerns have closed out their agricultural research, and pharmaceutical houses have likewise decreased their synthesis and screening operations.

For those research organizations that continue in the quest for new compounds with improved biological activity or that fulfill a market need, the risk is high, but the rewards are high also. It is little wonder that in an effort to decrease this risk, these concerns are devoting increasing attention to new tools that aid in the selection of compounds for further development.

The idea of correlating biological activity with chemical structure and physical properties of compounds is not new. Indeed, the beginnings may be traced back to the famous Meyer-Overton theory (*1*, *2*) where the bioactivity of anesthetics was correlated with lipid/water partition coefficients. In 1935 Meyer and Hemmi (*3*) disclosed that biological activity was again related to a partition coefficient. In 1962 Hansen (*4*) demonstrated the relationship between biological activity and the Hammett sigma constant. Each one of these approaches was applicable within a narrow area of structure and activity. In 1964 Hansch and Fujita (*5*) considerably broadened the scope and utility of structure–activity correlations with their introduction of an activity model and a two-parameter equation based on Hammett sigma constants and solubility parameters.

The Hansch-Fujita method, the Free-Wilson theory (described by Craig in Chapter 8), and the molecular orbital theory (described by Kier, Chapter 15) have greatly increased our understanding of the mode of action of many biological chemicals. It has also improved our ability to predict the activity of a variety of chemicals against plant and insect pests and certain pathogenic organisms.

One way to illustrate structure–activity correlations is to plot activity as a logarithmic function of a physical parameter. The resulting plot yields a curve where optimum activity lies at the summit of a hill. By knowing the effects of appropriate dependent variables on activity, one can learn how to reach the summit in the most efficient manner.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.pr001

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Adverse biological activity, such as toxicology, can also be correlated with physical and chemical parameters. Imagine such activity for a chosen family of compounds as being represented by a black hill. If the desired biological activity is correspondingly represented as a white hill, the peaks of the white and black hills will generally be displaced from each other. Hence, one may advantageously use correlation studies to approach as close to the top of the white hill as possible while still being as far removed from summit of the black hill as is feasible. This concept is depicted on the cover and jacket of this book.

Acknowledgment

I extend my sincere appreciation to Corwin Hansch of Pomona College for his particularly enthusiastic assistance in organizing this symposium. My thanks also goes to W. P. Purcell and T. R. Fukuto for their help and participation in the symposium.

I also extend appreciation to J. E. Johnson and W. Brock Neely of the Dow Chemical Co. for stimulating and maintaining my interest in this area of activity for many years. And last but not least, my sincere thanks go to A. F. Yapel here at 3M for his enthusiasm, drive, and hard work that made our participation in this symposium possible.

Literature Cited

1. Meyer, H. H., *Arch. Exptl. Pathol. Pharmakol.* Naunyn-Schmiedeberg's (1899) **42,** 109.
2. Overton, E., "Studien über die Narkose," Jena, Germany, 1901.
3. Meyer, K. H., Hemmi, H., *Biochem. Z.* (1935) **227,** 39.
4. Hansen, O., Rosenlund, *Acta Chem. Scand.* (1962) **16,** 1593.
5. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **86,** 1616.

St. Paul, Minn.
September 1972

J. W. VAN VALKENBURG

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.pr001

1

The Extrathermodynamic Structure-Activity Correlations

Background of the Hansch Approach

TOSHIO FUJITA

Department of Agricultural Chemistry, Kyoto University, Kyoto, Japan

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

The background supporting the versatility of the Hansch approach is that it is in fact an extrathermodynamic approach to drug action. The information on the structure–activity relationship of various kinds of drugs is expressed as equations which provide a convenient way for comparative studies with simpler and similar model reactions to elucidate the mechanism of overall drug action without microscopic knowledge of complex processes occurring in vivo.

In order for a drug to exhibit a certain biological effect, it must interact with a certain cellular component at the site of action. This component is often called a receptor. However, biological systems are composed of a number of heterogeneous phases, and the site at which a drug is administered is usually separated from the site of action. Thus, the drug must be transported through phase boundaries and undergo adsorption and desorption processes with proteins and membranes, as well as partitioning between different liquid phases, before it reaches the site of action. Moreover, the drug–receptor interaction at this site does not occur without perturbation by surrounding heterogeneous components such as water, serum protein, lipid particles, etc. Although the transport processes and the drug–receptor interaction are essentially physicochemical, they are far more complex than the homogeneous equilibria and rate processes of usual organic reactions. In this situation, it would rarely be possible to elucidate the mechanism of drug action if we were to insist upon only deterministic as well as microscopic models of individual stages of the transport and interaction processes.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

Recently, our Hansch approach has been widely accepted and recognized as a versatile way to understand drug action by analyzing the structure–activity relationship in various biological systems (*1, 2, 3*). This approach assumes that the physicochemical factors governing the transport and drug–receptor interaction can be factored into electronic, hydrophobic, and steric components. In general, one can consider that the variations in biological activity arising from structural modifications in congeneric drugs depend upon the concomitant changes in these physicochemical factors. The assumption is summarized as Equation 1 with parameters for the factored physicochemical properties, E, H, and S, respectively. In a series of congeners, the biological activity, BA, for a

$$\Delta \mathrm{BA} = f(\Delta E, \Delta H, \Delta S) \tag{1}$$

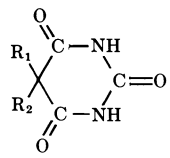
$$\mathrm{BA} = f(\Delta E, \Delta H, \Delta S) + \text{constant} \tag{2}$$

reference compound is a constant . Thus, Equation 2 holds for each member of the series. Usually, the value of BA is taken to be reciprocally proportional to the drug concentration, C, which causes a standard biological response such as EC_{50}, I_{50}, minimum inhibitory concentration, etc. on a molar basis. More practically, the free-energy related parameters are used such as log $1/C$ for BA, π or log P for ΔH, σ, σ^*, $\sigma^{\cdot}$, or $\Delta \log K_A$ for ΔE and E_s or E_s^c for ΔS. π is the hydrophobic substituent constant derived from the partition coefficient, P, which is usually determined with the use of 1-octanol/water system (*4, 5*), σ is the Hammett constant (*6*), $\sigma^{\cdot}$ is an electronic parameter for radical reactions (*7*), σ^* and E_s are the Taft's polar and steric constants for aliphatic systems (*8*), and E_s^c is the Hancock corrected steric constant (*9*). The most widely used equation in this approach is formulated as Equation 3, where $a(\geq 0)$, b, ρ, δ, and c are constants which are determined by the regression analysis using the least-squares method.

$$\log 1/C = -a(\log P)^2 + b \log P + \rho\sigma + \delta E_s + c \tag{3}$$

In this way it is possible to analyze how each of the physicochemical properties of the molecule is concerned with the drug action. Some examples are shown in Equations 4–11. In these equations, n is the number of points used in the regression, s is the standard deviation, and r is the correlation coefficient.

NADH Oxidation, Inhibition by Barbiturates (*3*)



$$\mathrm{pI}_{50} = 1.11 \log P + 1.24 \qquad n = 6 \quad s = 0.261 \quad r = 0.921 \tag{4}$$

Chymotrypsin, Inhibition by Miscellaneous Neutral Compounds (2)

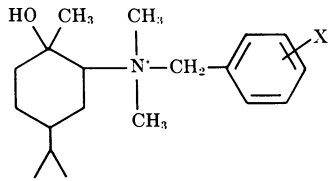
$$pK_I = 1.00 \log P - 2.60 \qquad n = 8 \quad s = 0.139 \quad r = 0.995 \qquad (5)$$

Cat, Inhibition of Salivary Secretion, Benzilic Acid Choline Esters (10, 11)

$$\log BA = -3.31\Sigma\sigma^* - 0.55\Sigma\pi + 2.30$$

$$13 \quad 0.159 \quad 0.971 \qquad (6)$$

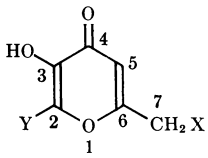
Alfalfa, Growth Inhibition, Benzyl Quaternary Ammonium Ions (11, 12)



$$pI_{50} = -1.02\sigma + 0.64\pi + 1.50$$

$$9 \quad 0.119 \quad 0.930 \qquad (7)$$

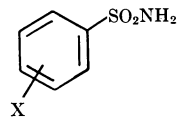
S. Aureus, Bacteriostatic Activity, Kojic Acid Analogs (pH 7) (13)



$$\log \frac{1}{C_n} = 0.97\,\Delta\log K_A + 1.49\pi(X) + 1.71$$

$$9 \quad 0.511 \quad 0.956 \qquad (8)$$

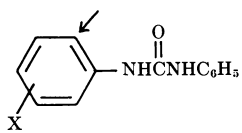
Carbonic Anhydrase, Inhibition, Benzenesulfonamides (14)



$$pK_I = 0.77\,\Delta\log K_A + 0.38\pi + 0.52$$

$$16 \quad 0.216 \quad 0.938 \qquad (9)$$

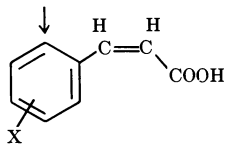
Tobacco Pith, Cytokinin Activity, Diphenylureas (15, 16)



$$\log 1/C = 1.10\sigma + 0.36\pi + 5.68$$

$$9 \quad 0.178 \quad 0.945 \qquad (10)$$

Wheat Coleoptile Cylinder,
Auxin Activity, cis-*Cinnamic Acids* (*15, 17*)



$$\log 1/C = -0.50(\log P)^2 + 2.45 \log P + 1.92\sigma + 1.77$$

8 0.290 0.815 (11)

In Equations 10 and 11 the σ values are those to the ortho position of the side chain.

In classical structure–activity studies, most of the attempts concentrated on correlating the activity with one of the molecular properties—*e.g.*, the carcinogenic activity of polynuclear aromatic compounds with their electronic structure (*18, 19*), the narcotic activity with lipophilicity (*20, 21*), the insecticidal activity of cyclodienes with their three-dimensional molecular silhouette (*22*), etc. Sometimes the activity correlated well with only one of the molecuar parameters. In our approach these are special cases where other physicochemical properties do not play critical roles in determining the variation in the activity within a set of congeners so that the coefficients defining these other properties are zero.

The parameterization of the hydrophobic character is one of the most important aspects of this approach. The π value of a certain substituent or a part of a molecule is nearly constant in closely related molecules and may be summed with other π values to calculate and predict unknown log P values of a number of molecules (*3, 4, 5*). Thus, the analyses can often be performed with the calculated log P values. One example for such an additive principle is the 1-octanol/H_2O partition coefficient of β-BHC (1*e*, 2*e*, 3*e*, 4*e*, 5*e*, 6*e*-hexachlorocyclohexane) which has been determined recently by Kurihara (*23*). The log P value, 3.78, is divided by 6 to estimate the π value of a >CH—Cl unit. The value,

$$\log P(C_6H_6Cl_6)/6 = 0.63, \qquad \pi_{calc.}(>CH—Cl) = 0.60$$

0.63, agrees very well with that calculated as $\Sigma\pi$, 0.60, using 0.41 for a ring carbon atom, 0.39 for an aliphatic chlorine atom, and −0.20 for a branching structure.

Since the π value is constitutive, the stereospecific nature of hydrophobic bonding for drug–receptor interactions can be delineated by regression analyses with the π values of substituents separately for each position of the congeners. Thus, the substituent effect on the emulsin hydrolysis of substituted phenylglucosides has been nicely delineated by analyzing kinetic constants separately for meta and para isomers. The meta substituents play no hydrophobic role in the enzyme–substrate complex formation (*24*).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

Another example is the case of bacteriostatic activity of kojic acid analogs which is analyzed in Equation 8 (*13*). In this equation, C_n is the minimum inhibitory concentration for the neutral molecule (*vide infra*), and $\pi(X)$ is the parameter for substituents at the 7 position. Neither the use of $\pi(X + Y)$ nor the addition of $\pi(Y)$ term improve the correlation. The substituents, Y, at the 2 position, sandwiched between hydroxyl and ether oxygen, might be surrounded by tightly hydrated water molecules so that they are unable to participate in the binding of the molecule with a hydrophobic surface of the receptor.

The assumption that the nonlinear dependence of drug activity on the hydrophobic character of the molecule is expressed by the quadratic terms of log P or π has been exemplified by a number of congeners where the log P values cover a sufficiently large range (*2, 3, 25*). Recently this has been proved theoretically with a kinetic model similar to those used in the multicompartmental analyses (*26*) as well as by a simple model based on probability concepts (*27*). The optimum value of log P, log P_o, obtained by setting the derivative $\partial \log 1/C/\partial \log P$ equal to zero, is a useful parameter for designing the most potent drug in a set of congeners as well as for illustrating the character of barriers through which drugs have to travel (*25, 28*).

From Equation 11 the log P_o value for the growth promotion of wheat coleoptile segments with cinnamic acids is calculated as 2.45. This value is very close to those obtained for the growth promotion of avena coleoptile segments with substituted phenoxyacetic and phenylacetic acids, the log P_o values being around $2 \sim 2.5$ (*29, 30*). Thus, the physicochemical character of barriers to reach the site of auxin action in plant tissues would be similar for different sets of compounds. Similar situations have been observed for the log P_o values of various sets of drugs inhibiting growth of gram-negative and gram-positive bacteria (*25*).

One question sometimes asked by biologists is why the drug action on such various levels of biological systems as enzyme preparations and subcellular organelles as well as whole plants and animals can be analyzed by a common procedure. For instance, the enzyme-inhibitor reaction seems a much simpler phenomenon than the action of cytokinins on plant tissues and, yet, they are illustrated in a similar manner as shown in Equations 9 and 10.

In fact, the enzyme-inhibitor interaction in itself is a chain of complex processes for the inhibitor molecule, including a number of desolvations, collisions with nonspecific sites on the enzyme protein, and resolvations before reaching the specific inhibition site. The complexity is not at all less than those considered for the transport and drug–receptor interaction processes of cytokinins. The situation is analogous to that a set of every rational number between zero and one corresponds in a one-to-one

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

fashion and thus is equivalent to a set of every integer between zero and infinity—*i.e.*, a "set" of barriers which inhibitor molecules must traverse to reach the inhibition site of the enzyme would be equivalent to a "set" of those for the cytokinin activity. The enzyme-inhibitor interaction processes could be a miniature model of the drug action in biological systems of the higher level. Thus, whichever may be the level of biological systems, essentially the same procedure can be applied for the drug action.

Another question which often arises is why such diverse biological effects of different physiological significance are rationalized by essentially the same type of equation where only physicochemical properties of drugs are considered. As a suitable approximation, any biological effect can be defined by two aspects. One is its specific physiological character such as photosynthetic inhibition, parasympatholytic action, bacteriostatic action, etc., which is determined by the biological test object and the specificity of the receptor. The other aspect is the magnitude of the effect which can be measured quantiatively. It can be further divided into sub-elements such as affinity and intrinsic activity as discussed by Ariëns (*31*).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

It is generally accepted that a drug initiates a chain of events which eventually leads to a specific biological effect but which does not involve the drug after it triggers the mechanism through a drug–receptor interaction. For example, sucrose tastes sweet, but the role of sucrose molecules is to stimulate the taste buds, and they do not participate in the process of sensory conduction as such.

The magnitude of the observable biological response is a direct reflection of the intensity of the chain of physiological events which, in turn, is determined by the degree of drug–receptor interaction. The degree of drug–receptor interaction is controlled by the physicochemical properties of the partners and the drug concentration at the site of action, which, in turn, is governed by the physicochemical transport process. Thus, as far as a specific biological effect is concerned, it is reasonable to consider that the magnitude of the response is expressed by a function of the physicochemical properties of the drug molecule. It is the magnitude of the biological effect which is used for the structure–activity analyses.

An interesting example which would illustrate the above situation is the case of DDT and its analogs although their structure–activity relationship has not been successfully analyzed by Equation 3. DDT is well known as an insecticide which kills by disrupting nerve conduction (*32*). DDT analogs where the *p,p'*-substituents on the benzene rings are changed are known to exhibit various degrees of activity (*33*). DDE, which is formed by dehydrochlorination of DDT, is not insecticidal. A

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

similar pattern of activity for this class of compounds has been observed for the inhibition of the cyclic photophosphorylation of barley (*34*). Thus, even if the observed physiological effect is distinctly different, the magnitude of activity seems to follow the same principle. For DDT and its analogs the main concern is to find a way to the interaction with receptor sites which is capable of matching their physicochemical properties regardless of whether the sites are located on the insect nerve or on the chloroplasts.

Even if the important role of physicochemical factors governing the magnitude of drug activity would become clearer, one might still be reluctant to accept this approach since the procedure is empirical rather than theoretical. With the use of regression analysis, it is possible to separate the relative significance of physicochemical factors in the overall action of a drug. Of course, one or a few of the overall processes could be expected to be critical in determining the drug activity. However, we cannot identify these particular processes, especially when for example the whole animal body is used to evaluate drug action. Thus, even if an equation which shows a significant role of the electronic factor is obtained for congeneric drugs, physical organic chemists may be distrustful of using the Hammett σ or the Taft σ^* constant for a reaction which cannot be identified. Even so, the separation of physicochemical factors playing roles in the overall drug activity is an important point from which to start a search for elucidating the mechanism of the drug action. Some examples are considered in the following sets of comparative correlations.

Substituted Phenyl Diethyl Phosphates

Rate of alkaline hydrolysis *in vitro* (*35*)

	n	s	r	
$\log k_{hyd.} = 1.35\sigma^- - 5.09$	7	0.238	0.955	(12)

Inhibition against flyhead acetylcholinesterase (*35, 36*)

$pI_{50} = 2.37\sigma^- + 4.38$	6	0.297	0.985	(13)

Toxicity against house fly (*1, 35*)

$-\log LD_{50} = 0.36 \log P + 2.65\sigma^- + 2.44$	8	0.206	0.990	(14)

The positive ρ values of these correlations for reactions of substituted phenyl phosphates may indicate a common feature in these reactions where a nucleophilic attack on phosphorus is a critical step. The magnitudes of the ρ values of Equations 13 and 14 are very close to each other but significantly larger than that of Equation 12. Thus, the electronic effect of the substituents on the insect mortality can be related mostly to an effect on the enzyme inhibition and plays only a minor role in the

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

transport step. However, besides the usual nucleophilic attack on phosphorus which occurs *in vitro,* other steroelectronic factors should be considered for the *in vivo* enzymatic reactions.

Substituted Phenyl N-Methyl Carbamates

Rate of alkaline hydrolysis *in vitro* (*37*)

	n	s	r	
$\log k_{[OH^-]} = 2.48\sigma^- + 3.03$	6	0.156	0.977	(15)

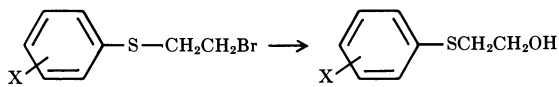
Flyhead acetylcholinesterase inhibition (*38, 39*)

$pI_{50} = 0.69\pi - 0.95\sigma^- + 1.19X + 3.50$	53	0.415	0.913	(16)

A more complex situation is observed with carbamate insecticides. Here, the sign of the ρ value on Equation 16 for the enzyme inhibition is opposite that in Equation 15 for the *in vitro* alkaline hydrolysis. In Equation 16, X is a position parameter which takes a value of 1 for meta substituents and 0 for para substituents. The enzyme inhibition has been considered to occur by carbamylation at the serine OH in the active site of acetylcholinesterase, and thus the mechanism should include a step for the nucleophilic attack of the serine OH on the carbamyl group. The negative sign of ρ and the significance of the X term in Equation 16 should indicate that, at least, the carbamylation of the enzyme is not the critical step, and quite different stereoelectronic conditions are required for this series of compounds from those for the *in vitro* nucleophilic hydrolysis and also from those for the enzyme–inhibitor interaction of the phosphate insecticides. Although the difference in the electronic effect of substituents on the enzyme inhibition has been recognized to exist between carbamates and phosphates, only by separating physicochemical effects would we know to what degree they are different and whether or not other effects may play roles.

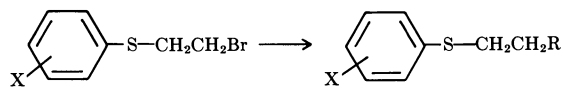
2-Bromoethylthiobenzenes

Hydrolysis *in vitro* (*40*)



	n	s	r	
$\log k(30°) = -1.23\sigma - 6.59$	8	0.048	0.984	(17)

Rate of alkylation *in vitro* (*41*)



$\log k(30°) = -1.98\sigma - 4.10$	12	0.073	0.994	(18)

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

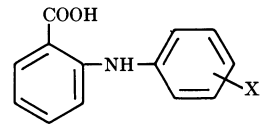
Toxicity against eggs of
Tetranychus telarius (L.) (*42*)

	n	*s*	*r*	
$-\log LC_{50} = -2.18\pi^2 + 1.69\pi - 1.45\sigma + 4.38$	8	0.164	0.968	(19)

As shown in Equation 19, the substituent effects on the ovicidal activity can be separated into hydrophobic and electronic factors. There is an optimal value for hydrophobicity for the ovicidal action of this set of compounds. That the σ value in Equation 19 is similar to those obtained for *in vitro* reactions would suggest that a common step is critical in biological as well as *in vitro* reactions. For the *in vitro* reactions, the rate-determining step is considered to be the formation of cyclic sufonium ions (*40, 41*).

Fenamic Acids

Uncoupling activity with
rat heart mitochondria (*43*)



	n	*s*	*r*	
$\log 1/C_i = 0.78\pi + 0.36\,\Delta\log K_A + 4.32$	11	0.200	0.939	(20)

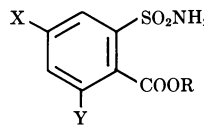
Anti-inflammatory activity against
antiserum induced rat edema (*43*)

$\log BA_i = 0.39\pi + 0.37\,\Delta\log K_A + 1.55$	12	0.106	0.867	(21)

As shown in Equations 20 and 21, the uncoupling activity and anti-inflammatory activity of fenamic acid analogs, including flufenamic and mefenamic acids, are very similar in their dependence on both hydrophobic and electronic effects of substituents. In these equations, the subscript, *i*, means that the drug activities are calculated on the basis of concentration of the ionized form. It has been suggested that these two biological effects, which are also observed in many other acidic anti-inflammatory drugs, have similar physicochemical mechanisms in the interaction with receptors (*44*).

Alkyl 2-Sulfamoylbenzoates

Mouse, antistrychnine activity (*45*)



$$\log 1/C = 0.21\,E_s(R) - 0.24\log P - 3.87\,\sigma^*(R) + 3.05$$

n	*s*	*r*	
9	0.058	0.959	(22)

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

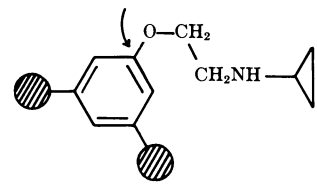
Mouse, antielectroshock activity (*45*)

$$\log 1/C = 0.40\,E_s(R) - 0.92\log P - 6.34\,\sigma^*(R) - 0.53\,\Sigma\sigma(X,Y) + 2.91 \qquad n = 12,\ s = 0.224,\ r = 0.860 \qquad (23)$$

Hamor and Lien have analyzed anticonvulsant activities of sulfamoylbenzoates. For the test of antistrychnine activity, compounds having the same aromatic substituents are used so that no $\Sigma\sigma(X,Y)$ term appears in Equation 22. They have suggested that the similarity of equations in terms of steric, hydrophobic, and electronic properties of substituents indicates a common anticonvulsant mechanism for the two biological effects of this set of compounds. They have also suggested that the mechanism of action of these drugs was quite different from those of barbiturates and other hypnotics where quite different structure–activity correlations of physicochemical significance have been obtained.

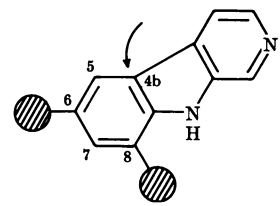
MAO Inhibitors

Rat liver MAO, phenoxyethylcyclopropylamines (*46*)



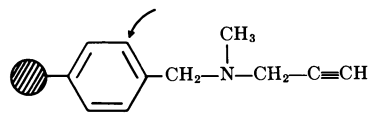
$$pI_{50} = 0.70\,E_s^{mm} + 1.64\sigma + 0.20\pi + 4.15 \qquad n = 18,\ s = 0.330,\ r = 0.945 \qquad (24)$$

Beef liver MAO, substituted β-carbolines (*47*, *48*)



$$pI_{50} = 0.61\,E_s^{6,8} + 0.72\sigma_{4b} + 0.52\pi + 3.03 \qquad 8\ \ 0.124\ \ 0.985 \qquad (25)$$

Rat liver MAO, pargyline derivatives (*47*, *49*)



$$pI_{50} = 0.79\,E_s^4 + 1.02\sigma_2 + 0.44\pi + 5.49 \qquad 10\ \ 0.273\ \ 0.942 \qquad (26)$$

Kutter and Hansch have analyzed the monoamine oxidase inhibition of phenoxyethylcyclopropylamines (*46*). Equation 24 shows that specific steric effects as well as electron-attracting and hydrophobic properties of substituents are responsible to the activity. Recently, it has been shown that the E_s value can be used as an index for intermolecular steric effects (*46*). In Equation 24, E_s^{mm} is the sum of E_s values of substituents at the two meta positions. The positive sign of the coefficient of this term means that the corresponding positions on the receptor site cannot accommodate larger substituents because of steric restraint.

Similar substituent effects are observed for the anti-MAO activity of β-carbolines and pargyline derivatives (*47*). In Equation 25, $E_s^{6,8}$ is the sum of values of the 6 and 8 substituents, and σ_{4b} is the σ value of substituents to the 4b position. In Equation 26, E_s^4 corresponds to the para substituent and σ_2 to the ortho position of the side chain. The coincidence in the stereoelectronic effects of substituents on the MAO inhibition is surprisingly good for these three classes of inhibitors. On the structural formulas shown above, the arrows indicate positions where the σ values are directed, and the shaded circles show sterically limited positions. Thus, the individual correlations are substantiated, and physicochemical significance is proved by similar biological correlations.

The examples above are only a few of numerous instances which have been found so far. In many cases it is possible to find clues to develop the studies on the mechanism of drug action by examining differences or common features among structure–activity and structure–reactivity relationships. Even if detailed microscopic mechanisms on the overall processes of drug action are not identified explicitly, explanations of substituent effects and sometimes information about critical process(es) determining the activity can be obtained by comparison with those of simpler and similar model reactions and/or drug actions with the use of free-energy related parameters. Thus, our procedure can be called an extrathermodynamic approach (*50*) to drug action.

Even more important, the equations are a convenient way to store and retrieve information on structure–activity relationships when combined with the use of a large computer. If the enormous amount of information accumulated so far is stored in the computer, one can sort out a set of equations according to their characteristic features such as the magnitude and sign of the coefficient of each terms and intercept. In this way the structure–activity correlations obtained for a certain series of congeneric drugs on various biological systems and for various series of drugs on a certain biological system or for various series of drugs on various biological systems can be classified and compared in terms of their physicochemical significance. As illustrated elegantly by Corwin

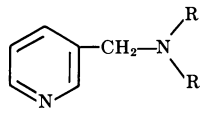
Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Hansch (Chap. 2), the computerized approach should be a serious step toward quantitative comparative pharmacodynamics.

While the versatility is well accepted, one should be cautious in practicing this approach. The lack of additivity of π and log P values has been found in molecules where intramolecular interactions are not negligible (*51, 52, 53*). The log P values of β-pyridylmethyldialkylamines have been determined as shown in Table I (*52*). The values calculated on the basis of $\Sigma\pi$, π(pyridine) = 0.65, $\pi(-CH_2-)$ = 0.5 and $\pi[N(CH_3)_2]$ = −0.35, are quite different from the corresponding experimental values. The lack of additivity should indicate peculiar intramolecular interactions depending upon the side chain structure. In such cases, it is essential to use the experimental log P values for structure–activity correlations.

Table I. Log P Values of β-Pyridylmethyldialkylamines



R		$-CH_3$	$-C_2H_5$	$-C_3H_7$	$-iso\text{-}C_3H_7$	$-C_4H_9$
log P	obsd.	0.49	1.01	1.46	2.27	1.47
	calcd.	0.80	1.80	2.30	1.90	2.80
	$\lvert\Delta\rvert$	0.31	0.79	0.84	0.37	1.33

Although the lack of additivity is, in a way, a drawback to the simplicity and versatility of this approach, the difference between calculated and experimental log P values can sometimes give important information on the conformation of the molecule in the aqueous phase (*51, 54*). The π values of polar substituents obtained in the alkyl derivatives are higher than corresponding values calculated from the arylalkyl derivatives. The differences (nearly constant around 0.6) are thought to arise from intramolecular hydrophobic bonding of the side chain with the aromatic ring in arylalkyl compounds which causes a bending conformation in the aqueous phase. Detailed studies on the nature of log P and π parameters are discussed by several authors in subsequent chapters.

It is important to determine whether or not each substituent parameter is an independent variable. If a considerable correlation exists between two parameters in a set of compounds, it is often difficult to evaluate the quality of structure–activity correlations. The data in Table II are for the muscarine-like activity of acetylcholine analogs measured by the blood pressure descent of cats (*55*). The effect of cationic head structure has been analyzed using physicochemical parameters of the N substituents (*11*).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

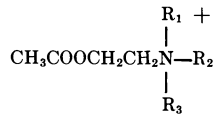
	n	s	r	
$\log BA = -5.18\,\Sigma\sigma^* + 4.22\,\Sigma E_s^c - 0.063$	7	0.490	0.954	(27)
$\log BA = 26.62\,\Sigma\pi + 37.89\,\Sigma E_s^c - 0.063$	7	0.490	0.954	(28)
$\log BA = -5.83\,\Sigma\sigma^* - 3.34\,\Sigma\pi - 0.063$	7	0.490	0.954	(29)

Equations 27–29 with two parameters seem to show reasonable correlation. Although Equation 28 would be unacceptable because of unusually large parameter coefficients, the correlations are completely equivalent statistically. As far as compounds used for the analyses are concerned, it is impossible to choose the two significant parameters. In this case, besides a significant correlation between parameters π and E_s, there are mutual relationships among three parameters. Each parameter is expressed as a linear combination of the other two and is not separated from others.

To obtain meaningful correlations, substituents must be selected so that the substituent parameters included in a set of congeners are as separated as possible. In this example, not only the hydrogen, methyl, and ethyl, whose parameters vary in parallel, but groups such as isopropyl ($\sigma^* = -0.19$, $\pi = 1.20$, $E_s^c = -0.47$), *tert*-butyl ($\sigma^* = -0.30$, $\pi = 1.60$, $E_s^c = -1.54$), or trifluoroethyl ($\sigma^* = 0.92$, $\pi = 1.48$, $E_s^c = -0.79$) could be used as the N substituents. The situation has been reviewed recently by Craig (*56*).

Sometimes the use of quantum chemically determined indices (*1, 24, 57*) or oxidation–reduction potentials (*58*) instead of the Hammett-Taft constants, the use of chromatographically obtained hydrophobicity parameter, ΔR_m, (*5, 59*) instead of π or log P, or the addition of other variables such as those for hydrogen bonding (*60*), dipole moments (*61*),

Table II. Muscarine-like Activity of Acetylcholine Analogs



Substituents			*Parameters*			
R_1	R_2	R_3	$\Sigma\sigma^*$	$\Sigma\pi$	ΣE_s^c	*log BA*[a]
Me	Me	Me	0.00	0.00	0.00	0.00
Me	Me	H	0.49	−0.50	0.32	−1.70
Me	H	H	0.98	−1.00	0.64	−2.70
H	H	H	1.47	−1.50	0.96	−3.30
Me	Me	Et	−0.10	0.50	−0.38	−0.48
Me	Et	Et	−0.20	1.00	−0.76	−2.60
Et	Et	Et	−0.30	1.50	−1.14	−3.30

[a] BA is the activity relative to acetylcholine.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

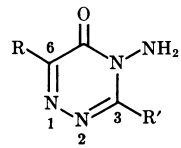
bond polarizations (*62*), and position parameters (*39*) in Equation 3 can be justified for the improved correlation.

In any case, it is advisable not to use too many parameters in Equation 3. Unless the number of compounds used in the regression is large so that the degrees of freedom are sufficient, a statistically significant correlation would not be obtained. For example, the depolarizing activity of five quaternary ammonium ions against electric eel electroplax (*63*) is expressed by Equation 30 with the sum of the substituent constants of four N-substituents (*11*).

$$\log A = -4.51\Sigma\sigma^* + 4.21\Sigma\pi + 6.56\,\Sigma E_s^c + 3.14 \qquad n = 5 \quad s = 0.162 \quad r = 0.992 \qquad (30)$$

The correlation coefficient is very high, and the standard deviation seems rather low. However, an F test reveals that the correlation is not justified at the 0.90 level of probability. In Equation 30, the number of degrees of freedom, equal to one, is so low that the F value, which is a function of the number, becomes rather small. ($F_{3,1} = 26.6$, $F_{3,1,0,10} = 53.6$) In general, the larger the number of degrees of freedom, the more significantly reduced is the variance in the activity data. When a statistically significant correlation is obtained with fewer parameters and yet an improved correlation is explored by adding another parameter, the level of confidence at which the additional parameter is justified should be examined with F tests.

It is not enough for a multiple parameter equation just to "work" and to give an improved correlation as Higuchi and Davis have emphasized recently (*64*). The correlation must have a physicochemical significance. Recently, Equation 31 has been presented by Büchel and his co-workers for the inhibition of the Hill reaction with 1,2,4-triazinone herbicides (*65*).



$$\mathrm{pI}_{50} = 5.23 - 16.57\,\delta + 45.04\,\delta^2 + 2.11\,\Delta R_M(3) - 3.36\,\Delta R_M(3)^2 + 2.22\,\Delta R_M(6) - 0.50\,\Delta R_M(6)^2 \qquad n = 28 \quad s = 0.418 \quad r = 0.938 \qquad (31)$$

In this equation, δ is a differential parameter developed from Sephadex and polyamide thin-layer chromatography, and $\Delta R_M(3)$ and $\Delta R_M(6)$ are the values for the substituents at the 3 and 6 positions, respectively. The number of compounds is sufficiently large so that the correlation could be accepted statistically. Although Equation 31 can be used to predict

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

activities of untested compounds in this particular set of compounds, it is difficult to accept its physicochemical significance on a general mode of action of Hill reaction inhibitors. The more parameters used for the correlation, the more difficult it would be to determine comparative correlations and the less attractive will be the physicochemical elucidation.

For a set of ionizable drugs where large changes in the degrees of ionization are observed, the apparent activity exhibited by a drug is not a suitable index to be analyzed. Unless the effect of ionization is separated from other physicochemical effects, the correlation would not have a physicochemical significance. The apparent activity, log $1/C$ is, in fact, log $1/(C_{\text{ion}} + C_{\text{neutral}})$, which is not a free-energy related index for the magnitude of activity either of the neutral or ionized molecule or both. With a relativistic approach to this problem, we plan to use Equation 32 and 33 for the ionizable drugs (*66, 67*).

$$\log 1/C + \log (K_A + [\text{H}^+])/[\text{H}^+] = -a(\log P)^2 + b \log P + \rho\sigma + \delta E_s + c \quad (32)$$

$$\log 1/C + \log (K_A + [\text{H}^+])/K_A = -a(\log P)^2 + b \log P + \rho'\sigma + \delta E_s + c' \quad (33)$$

From the bacteriostatic activity data of kojic acid derivatives determined at pH 6, 7, and 8 with *s. aureus*, we have derived the following equations (*13*).

For the apparent activity:

pH 6: $$\log 1/C = 0.863\ \Delta\log K_A + 1.587\ \pi(X) + 1.571 \quad n = 9 \quad s = 0.585 \quad r = 0.946 \quad (34)$$

pH 7: $$\log 1/C = 0.568\ \Delta\log K_A + 1.456\ \pi(X) + 1.632 \quad n = 9 \quad s = 0.516 \quad r = 0.944 \quad (35)$$

pH 8: $$\log 1/C = 0.358\ \Delta\log K_A + 1.205\ \pi(X) + 1.600 \quad n = 9 \quad s = 0.476 \quad r = 0.930 \quad (36)$$

For the neutral molecule:

pH 6: $$\log 1/C + \log (K_A + [H^+])/[\text{H}^+] = 0.959\ \Delta\log K_A + 1.601\ \pi(X) + 1.580 \quad n = 9 \quad s = 0.586 \quad r = 0.949 \quad (37)$$

pH 7: $$\log 1/C + \log (K_A + [\text{H}^+])/[\text{H}^+] = 0.970\ \Delta\log K_A + 1.487\ \pi(X) + 1.708 \quad n = 9 \quad s = 0.511 \quad r = 0.956 \quad (8)$$

pH 8: $$\log 1/C + \log (K_A + [\text{H}^+])/[\text{H}^+] = 1.132\ \Delta\log K_A + 1.235\ \pi(X) + 1.974 \quad n = 9 \quad s = 0.460 \quad r = 0.960 \quad (38)$$

For the ionized form:

pH 6: $$\log 1/C + \log (K_A + [\text{H}^+])/K_A = 1.606\ \pi(X) + 3.517 \quad n = 9 \quad s = 0.544 \quad r = 0.938 \quad (39)$$

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

pH 7: $$\log 1/C + \log (K_A + [H^+])/K_A = 1.491\,\pi(X) + 2.651 \quad n = 9 \quad s = 0.474 \quad r = 0.944 \quad (40)$$

pH 8: $$\log 1/C + \log (K_A + [H^+])/K_A = 1.215\,\pi(X) + 2.009 \quad n = 9 \quad s = 0.449 \quad r = 0.927 \quad (41)$$

Combined equations, pH 6, 7, and 8:

$$\log 1/C + \log (K_A + [H^+])/[H^+] = 1.020\,\Delta\log K_A + 1.439\pi + 1.754 \quad n = 27 \quad s = 0.509 \quad r = 0.943 \quad (42)$$

$$\log 1/C + \log (K_A + [H^+])/K_A = 1.437\pi + 2.726 \quad n = 27 \quad s = 0.832 \quad r = 0.823 \quad (43)$$

In these equations, the $\pi(X)$ is that of the substituents, such as hydrogen, halogen, and hydroxyl, at the 7-position. Neither the use of $\pi(X + Y)$ nor the addition of $\pi(Y)$ term improve the correlations as described before for Equation 8. The correlations are almost equivalent, qualitatively speaking, although those for the activity of ionized form are slightly better than the others as far as they are considered separately. Thus, it might seem of no great value to consider the effect of ionization. However, if we combine Equations 8, 37, and 38, Equation 42 is obtained for the activity of the neutral molecule at three different pH's with the use of 27 data points. Equation 42 shows much better correlation than Equation 43, which is derived in a similar manner for the activity of the ionized form. Since the bacterial growth is not affected much by pH changes from 6 to 8, the form of the kojic acid derivatives responsible for the bacteriostatic activity is likely to be the neutral molecule, and the ionized form seems to play only a minor role in the activity, if any. Equations such as 34, 35, and 36, which might be only of use to indicate a general trend, are, in effect, unable to elucidate the pH dependence of activity and to give any information about the active form.

It should be noticed that from the activity data measured at a certain fixed pH, the correlations of equivalent quality are obtained both for the neutral and ionized forms. There are interrelations *a priori* between Equations 32 and 33 such as $\rho - \rho' = \rho_A$ and $c - c' = \text{pH} - \text{p}K_A^{\text{std}}$, where ρ_A is the Hammett reaction constant for the ionization equilibrium and $\text{p}K_A^{\text{std}}$ is the value of a standard compound. Thus, it is only possible to predict the molecular form responsible for the activity by comparing Equations 32 and 33 derived from data obtained at various pH. Moreover, the optimum value of log K_A, log $K_A°$, for the apparent activity of ionizable congeners can be derived by setting the derivative of either Equation 32 or 33 equal to zero as shown in Equations 44 and 45.

$$\frac{\partial \log 1/C}{\partial \sigma} = \rho - \frac{K_A}{K_A + [H^+]}\rho_A = \rho' - \frac{[H^+]}{K_A + [H^+]}\rho_A = 0 \quad (44)$$

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

$$\log K_A^\circ = \log\left(-\frac{\rho}{\rho'}\right) - \mathrm{pH} \qquad (45)$$

Similar situations have been discussed for substituted phenols (*66*) and sulfa drugs (*67*) in detail.

When structural modifications are made at various positions in a set of complex drug molecules, this approach combined with the Free-Wilson mathematical model would be a suitable tool (*68*). If the mathematical contributions of substituents and parts of molecules are additive for a certain biological effect of a series of drugs, the contribution values of substituents at each position can be analyzed further with our approach. Examples are discussed by Paul Craig in Chapter 8.

When drug biotransformations and excretions occur before reaching the site of action and common mechanisms are considered for a set of drugs, the rate constants determined by the multicompartmental analysis or its counterparts could be analyzed so that the modes of metabolism, distribution, and excretion are elucidated physicochemically (*69, 70, 71*). However, for the cases where specific biotransformations for certain members or each member of congeneric drugs are involved—*e.g.*, for the plant-growth regulating activity of gibberellins (*72*)—the structure–activity relationship could not be successfully analyzed by Equation 3 or its modifications.

From the line of reasoning that any biological effect of drugs should be a result of chemical and/or physicochemical perturbation on the biological system, the relationship between structure and activity observed for any kind of drug should be analyzable on the basis of physicochemical parameters. Although there are examples where Equation 3 or its modification fails in analyzing the correlation, this does not obviate the physicochemical mechanism behind the drug action. We hope that this approach, which is still in an early stage, will achieve further developments and stimulate other related fields of science so that the mechanisms of the drug action can be understood comprehensively.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

Acknowledgment

The author thanks Corwin Hansch for invaluable discussions since our first project on this approach began in 1961.

Literature Cited

1. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **86**, 1616.
2. Hansch, C., *Accounts Chem. Res.* (1969) **2**, 232.
3. Hansch, "Drug Design," E. J. Ariëns, E., Vol. 1, p. 271, Academic, New York, 1971.

4. Fujita, T., Iwasa, J., Hansch, C., *J. Amer. Chem. Soc.* (1964) **86,** 5175.
5. Iwasa, J., Fujita, T., Hansch, C., *J. Med. Chem.* (1965) **8,** 150.
6. Hammett, L. P., "Physical Organic Chemistry," 2nd ed., p. 347, McGraw Hill, New York, 1970.
7. Hansch, C., *J. Med. Chem.* (1968) **11,** 920.
8. Taft, R. W., Jr., "Steric Effects in Organic Chemistry," p. 556, M. S. Newman, Ed., Wiley, New York, 1956.
9. Hancock, C. K., Meyers, E. A., Yager, B. J., *J. Amer. Chem. Soc.* (1961) **83,** 4211.
10. Ford-Moore, A. H., Ing, H. R., *J. Chem. Soc.* **1947,** 55.
11. Fujita, T., Nishioka, T., Takayama, C., Nakajima, M., unpublished data.
12. Newhall, F. W., *J. Agr. Food Chem.* (1971) **19,** 294.
13. Ichimoto, I., Kotani, T., Tatsumi, C., Fujita, T., "Abstracts of Papers," Annual National Meeting, Agricultural Chemical Society of Japan, Fukuoka, 1970, p. 110.
14. Kakeya, N., Yata, N., Kamada, A., Aoki, M., *Chem. Pharm. Bull.* (1969) **17,** 2558.
15. Fujita, T., *Chem. Regulation Plants (Tokyo)* (1970) **5,** 124.
16. Bruce, M. I., Zwar, J. A., *Proc. Roy. Soc.* (1966) **B165,** 245.
17. Ugochukwu, E. N., Wain, R. L., *Ann. Appl. Biol.* (1968) **61,** 121.
18. Nagata, C., Yonezawa, T., Fukui, K., Tgashira, Y., *Cancer Res.* (1955) **15,** 233.
19. Pullman, A., Pullman, B., *Adv. Cancer Res.* (1955) **3,** 117.
20. Meyer, K. H., Hemmi, H., *Biochem. Z.* (1935) **277,** 39.
21. Overton, E., *Z. Physiol. Chem.* (1897) **22,** 189.
22. Soloway, S. B., *Adv. Pest Control Res.* (1965) **6,** 85.
23. Kurihara, N., private communication (1971).
24. Hansch, C., Deutsch, E. W., Smith, R. N., *J. Amer. Chem. Soc.* (1965) **87,** 2738.
25. Lien, E. J., Hansch, C., Anderson, S. M., *J. Med. Chem.* (1968) **11,** 430.
26. Penniston, J. T., Beckett, L., Bentley, D. L., Hansch, C., *Mol. Pharmacol.* (1969) **5,** 333.
27. McFarland, J. M., *J. Med. Chem.* (1970) **13,** 1192.
28. Hansch, C., *Il Farmaco* (1968) **23,** 293.
29. Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., Streich, M., *J. Amer. Chem. Soc.* (1963) **85,** 2817.
30. Muir, R. M., Fujita, T., Hansch, C., *Plant Physiol.* (1967) **42,** 1519.
31. Ariens, E. J., "Molecular Pharmacology," Vol. 1, p. 176, Academic, New York, 1964.
32. Yamasaki, T., Narahashi, T., *Botyu-Kagaku* (1957) **22,** 296.
33. Metcalf, R. L., "Organic Insecticides," p. 135, Wiley, New York, 1955.
34. Upshall, D. G., Goodwin, T. W., *J. Sci. Food Agr.* (1964) **15,** 846.
35. Fukuto, T. R., Metcalf, R. L., *J. Agr. Food Chem.* (1956) **4,** 930.
36. Hansch, C., Deutsch, E. W., *Biochim. Biophys. Acta* (1966) **112,** 381.
37. O'Brien, R. D., Hilton, B. D., Gilmour, L., *Mol. Pharmacol.* (1966) **2,** 593.
38. Metcalf, R. L., Fukuto, T. R., *J. Agr. Food Chem.* (1965) **13,** 220.
39. Hansch, C., "Biochemical Toxicology of Insecticides," R. D. O'Brien and I. Yamamoto, Eds., p. 33, Academic, New York, 1970.
40. Ishida, S., Ida, M., *Agr. Biol. Chem. (Tokyo)* (1967) **31,** 410.
41. Ishida, S., Yamada, O., *Agr. Biol. Chem. (Tokyo)* (1967) **31,** 417.
42. Ishida, S., *Agr. Biol. Chem. (Tokyo)* (1966) **30,** 800.
43. Muraoka, S., Terada, H., Fujita, T., "Abstracts of Papers," 91st National Meeting, Pharmaceutical Society of Japan, Fukuoka, 1971, p. 412.
44. Whitehouse, M. W., "Pharmaceutical Chemistry—2," p. 39, Butterworth, London, 1969.
45. Hamor, G. H., Lien, E. J., *Il Farmaco* (1969) **24,** 704.
46. Kutter, E., Hansch, C., *J. Med. Chem.* (1969) **12,** 647.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch001

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

47. Fujita, T., Ban, T., unpublished data.
48. Ho, B. T., Li, Ko-Chin, Walker, K. E., Tansey, L. W., Kralik, P. M., McIsaac, W. M., *J. Pharm. Sci.* (1970) **59,** 1445.
49. Swett, L. R., Martin, W. B., Taylor, J. D., Everett, G. M., Wykes, A. A., Gladish, Y. G., *Ann. N.Y. Acad. Sci.* (1963) **107,** 891.
50. Leffler, J. E., Grunwald, E., "Rates and Equilibria of Organic Reactions," p. 128, Wiley, New York, 1963.
51. Hansch, C., Anderson, S. M., *J. Org. Chem.* (1967) **32,** 2583.
52. Fujita, T., Nakajima, M., Soeda, Y., Yamamoto, I., *Pesticide Biochem. Physiol.* (1971) **1,** 151.
53. Currie, D. J., Lough, C. E., Silver, R. F., Holmes, H. L., *Can. J. Chem.* (1966) **44,** 1035.
54. Leo, A., Hansch, C., *J. Org. Chem* .(1971) **36,** 1539.
55. Holton, P., Ing, H. R., *Brit. J. Pharmacol.* (1949) **4,** 190.
56. Craig, P. N., *J. Med. Chem.* (1971) **14,** 680.
57. Neely, W. B., White, H. C., Rudzik, A., *J. Pharm. Sci.* (1968) **57,** 1176.
58. Kutter, E., Machleidt, H., Reuter, W., Wildfeuer, A., 161st National Meeting, ACS, Los Angeles, Calif., 1971.
59. Boyce, C. B. C., Milborrow, B. V., *Nature* (1967) **208,** 2275.
60. Hansch, C., Lien, E. J., *Biochem. Pharmacol.* (1968) **17,** 709.
61. Clayton, J. M., Purcell, W. P., *J. Med. Chem.* (1969) **12,** 1087.
62. Hansch, C., Coats, E., *J. Pharm. Sci.* (1970) **59,** 731.
63. Podleski, T. R., *Biochem. Pharmacol.* (1969) **18,** 211.
64. Higuchi, T., Davis, S. S., *J. Pharm. Sci.* (1970) **59,** 1376.
65. Draber, W., Buchel, K. H., Dickore, K., 2nd International Congress of Pesticide Chemistry, Israel, 1971.
66. Fujita, T., *J. Med. Chem.* (1966) **9,** 797.
67. Fujita, T., Hansch, C., *J. Med. Chem.* (1967) **10,** 991.
68. Fujita, T., Ban, T., *J. Med. Chem.* (1971) **14,** 148.
69. Lien, E. J., Hansch, C., *J. Pharm. Sci.* (1968) **57,** 1027.
70. Lien, E. J., *Drug Intelligence Clin. Pharm.* (1970) **4,** 7.
71. Lien, E. J., Koda, R. T., Tong, G. L., *Drug Intelligence Clin. Pharm.* (1971) **5,** 38.
72. Crozier, A., Kuo, C. C., Durley, R. C., Pharis, R. P., *Can. J. Bot.* (1970) **48,** 867.

RECEIVED June 17, 1971.

2

A Computerized Approach to Quantitative Biochemical Structure–Activity Relationships

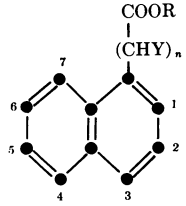
CORWIN HANSCH

Pomona College, Claremont, Calif. 91711

A general approach to the computerization of quantitative biomedical–chemical structure–activity studies is discussed. There are two main problems to consider. One is the formulation of quantitative relationships using physicochemical parameters and regression analysis. As such equations are derived, the problem of organization of the mass of data must be solved. The most suitable, relatively inexpensive method for dealing with the structures of organic compounds via *computers is the Wiswesser Line Notation method. This notation and the proper computer program can be of great help in comparative pharmacodynamics.*

Biochemical structure–activity studies in medicinal chemistry, agricultural chemistry, and enzymology continue to produce a flood of new papers each year which, when placed on the mountain of such research published over the last 70 years, defies organization. In certain areas such as antimalarials, anticancer drugs, and anticholinesterase compounds, so many molecules have been made and tested that no single mind can encompass the data much less integrate it into meaningful structure–activity patterns.

The awful job facing the drug designer once an active molecule has been uncovered can be illustrated with the general formula I.



I

Assume that a member of the above family has been found to have some desirable activity. In planning a drug modification study, let us consider the following set of functions for substitution at the seven ring positions of naphthalene.

Table I. Electronic and Hydrophobic Substituent Constants

Function	π^a	σ_p	*Function*	π^a	σ_p
H	0.00	0.00	OBu	1.48	−0.27
F	0.14	0.06	OC_6H_5	2.08	−0.03
Cl	0.71	0.23	$OCONH_2$	−1.05	
Br	0.86	0.23	$OCOCH_3$	−0.64	0.31
I	1.12	0.18	SH	0.39	0.15
Me	0.50	−0.17	SMe	0.61	0.00
Et	1.00	−0.15	SEt	1.11	0.00
Pr	1.50	−0.13	SF_5	1.23	0.68
Amyl	2.50	−0.13	SO_2NH_2	−1.82	0.57
Hexyl	3.00	−0.13	SO_2CH_3	−1.63	0.72
i-Pr	1.30	−0.15	SCF_3	1.44	0.50
i-Bu	1.80	−0.13	SO_2CF_3	0.55	0.93
t-Bu	1.68	−0.20	SC_6H_5	2.34	
i-Amyl	2.30	−0.13	NH_2	−1.23	−0.66
Cyclopropyl	1.20	−0.21	$NHCH_3$	−0.47	0.84
Cyclohexyl	2.51[b]	−0.15	$N(CH_3)_2$	0.18	−0.83
C_6H_5	2.13	−0.01	$N(Et)_2$	2.18	−0.83
$CH_2C_6H_5$	2.01	−0.09	$N(Bu)_2$	4.18	−0.83
C≡CH	0.40	0.23	NHC_6H_5	1,37	−0.40
CF_3	0.88	0.54	$NHCOCH_3$	−0.97	0.00
CCl_3	0.79	0.33	$NHCOC_6H_5$	0.49	−0.25
CN	−0.57	0.66	$NHCONH_2$	−1.31	0.24
OH	−0.67	−0.37	$N=NC_6H_5$	1.69	0.39
OMe	−0.02	−0.27	NO_2	−0.28	0.78
OEt	0.48	−0.27	$Si(CH_3)_3$	2.59	−0.07

[a] π-values are from the benzene system except where noted.
[b] These values are from the phenoxyacetic acid system.

The set of 50 functions of Table I is by no means extensive. However, it does provide some examples for consideration. For instance, no heterocyclic functions have been included, nor have we considered varying functions on the functions—*e.g.*, nitrophenyl, chlorophenyl, etc. Anyone could quickly extend the list to several hundred. Just to make all of the monofunctional derivatives at each position would mean 7 × 50 or 350 possible structures. If we consider polyfunctional derivatives, the numbers increase rapidly according to the formula X^N where X is the number of functions and N is the number of positions. The formula holds only as long as no elements of symmetry are introduced. The following calculations illustrate the awesome possibilities for a set of only 50 functions for substitutions in 2, 3, 4, 6, and 7 of the possible ring positions.

$$50^2 = 2500$$
$$50^3 = 125{,}000$$
$$50^4 = 6{,}250{,}000$$
$$50^6 = 312{,}500{,}000$$
$$50^7 = 15{,}625{,}000{,}000$$

Of course, to this must be added the possibilities resulting from variations in Y, n, and R. Take the case where we vary substituents at only four positions on the ring but, in addition, we make 10 changes in Y, 10 in R, and four in n. This leads to: $50^4 \times 10 \times 10 \times 4 = 2{,}500{,}000{,}000$. With the relatively simple limits set above, the chemist faces a staggering problem of selection. The problem becomes much worse if we entertain the possibility of introducing 1, 2, or 3 heteroatoms into the ring. Having so little in the way of theory for guidance, the simple way out is usually followed, and one makes the compounds which are easiest to synthesize. While the cost of synthesis is an aspect of drug research which cannot be taken lightly, with the great advances in modern synthetic techniques quite complex structures can be made if there is good reason to do so. Hopefully, computer techniques can be developed to help with the decision making in drug design as well as organization of the literature. The two major aspects of the problem are diagrammed in Figure 1.

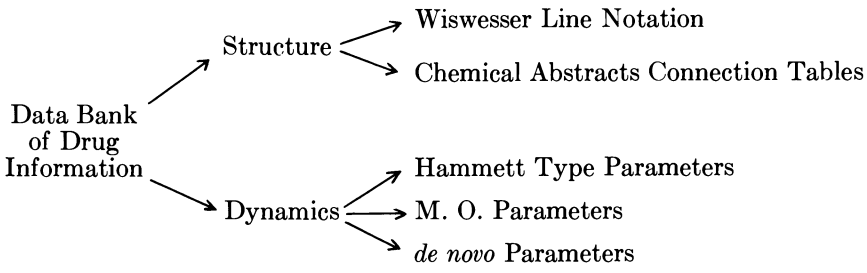
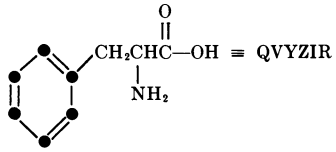


Figure 1. Model for computerization of structure–activity relationships

The computerization of organic structures has received considerable attention. Two general approaches represented by the Wiswesser line notation and connection table storage have received the most attention. Neither system has reached a state of perfection. Both have advantages and disadvantages. The Wiswesser (*1*) system is much less expensive in terms of computer storage space and search time. For these reasons we have elected to experiment with this method for organizing drug formulas. While proficiency in writing this language requires learning a large number of rules, learning to read the notation is relatively simple. The formula for phenylalanine is written as follows:



The two-dimensional structural formula reduces to a linear set of six symbols easily punched on an IBM card. Q represents OH, V is the symbol for —C(=O)—, Y indicates a branched chain to which an amino function (Z) and a CH_2 group (*1*) are attached. R stands for the benzene ring attached to the methylene function. Not only does such a formula require little storage space, but it makes substructure searching relatively easy. For example, one can query the data bank for all compounds having the grouping ZVO (*i.e.*, $H_2NC(=O)$–).

$$\text{System}_i + \text{X}_j\text{–drug} \rightarrow \text{Perturbation}_{ij} \quad (1)$$

$$\text{Perturbation}_{ij} = f(\text{Physicochemical parameters of X}_j) \quad (2)$$

Parameters

Hydrophobic: log P or π

Electronic: σ, σ^+, σ^-, σ^*, M.O., pK_a

Steric: E_s, molar volume, molecular refractivity

For a definition of these parameters, *see* Ref. *3*.

Although it is the dynamics of structure–activity studies with which we are most concerned at present, structure and dynamics must be well integrated. The above scheme represents our present model for computerizing the dynamic aspects of structural modification. The term system refers to the test system used, be it a mouse, an organelle, or an isolated enzyme. Equation 1 is like any chemical equation in that a given system + a reactant yields a product. However, in drug research one does not isolate the product; instead, it is characterized by an observed standard response (Perturbation$_{ij}$). In Equation 1 it is assumed that all of the drugs X$_j$–drug are acting in the same way by the same mechanism in producing the observed perturbation. Perturbation$_{ij}$ is probably best defined in terms of log $1/C$ where C is the molar concentration of drug producing the standard response in a fixed unit of time. One hopes to rationalize the perturbation with one or more independent variables

Table II. Log 1/C

Type of Compound	*Biological Activity*
Alcohols and ketones	I_{50} indophenol oxidation by rabbit kidney
ROH	Drop in black lipid membrane resistance from 10^8 to 10^6 ohm/cm^2
ROH	−10 mv change in rest potential of lobster axon
ROH	−5 mv change in rest potential of lobster axon
Miscellaneous	100% inhibition frog heart
Miscellaneous	I_{50} red blood cell oxygen consumption
ROH	Inhibition bacterial luminescence
ROH	Minimum inhibition concentration frog sciatic nerve
ROH	Narcosis of goldfish
ROH	I_{50} tortoise heart
Miscellaneous	Colchicine-like mitosis in onion root tip
ROH	Narcosis of barnacle larvae
ROH	Narcosis of 2.5-day tadpoles
ROH	Narcosis of 12-day tadpoles
ROH	I_{50} guinea pig ileum
Miscellaneous	100% inhibition tadpole movement
ROH	Narcosis of 83-day tadpoles

(Equation 2). Some of the commonly used variables are indicated in the scheme above. At this very early stage in the development of mathematical structure–activity work a complete set of well understood parameters has not been developed; however, the set above has at least allowed us to start serious work. The greatest lack at present is for a suitable numerical way to define the geometry of organic compounds—*i.e.*, one that does not require an inordinate number of terms.

The development of such extrathermodynamic equations places structure–activity relationships in a form easily stored and manipulated by computers. Many such examples have been published (*2*, *3*) and, at present, we are experimenting with a group of about 1000 such equations which correlate about 15,000 organic structures. Correlation equations now cover systems ranging from simple proteins (*4*) to pure enzymes (*5*), antibodies (*6*), organelles (*7*), and whole animals (*8*). As the data bank of such equations grows, specific information on many compounds accrues. To gain the maximum use of this information, the two systems of Figure 1 must be completely interactive. The following are examples of typical queries one would like to make of such a system *via* a remote terminal in the drug designer's laboratory.

(1) List all drugs containing the function $-\overset{\overset{O}{\|}}{S}CNH_2$ in order first of increasing log $1/C$ and then in order of log P. List the system in which each is active and the type of action measured.

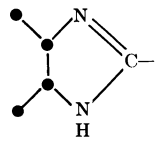
$= a \log P + b$

a	b	n	r	s	*Equation*
0.90±.11	−0.55±.06	8	0.993	0.064	3
1.26±.24	−0.52±.40	7	0.986	0.248	4
0.90±.22	−0.23±.18	5	0.991	0.112	5
0.86±.16	−0.09±.13	5	0.995	0.082	6
0.93±.08	0.11±.12	28	0.975	0.182	7
0.92±.12	0.12±.15	14	0.977	0.214	8
1.17±.08	0.22±.12	8	0.998	0.100	9
1.05±.10	0.27±.20	8	0.995	0.141	10
1.15±.20	0.34±.11	8	0.985	0.106	11
0.99±.20	0.52±.13	10	0.972	0.136	12
0.96±.14	0.52±.23	19	0.963	0.340	13
1.05±.09	0.57±.11	14	0.991	0.140	14
1.31±.11	0.59±.17	8	0.997	0.141	15
1.24±.08	0.61±.13	8	0.998	0.109	16
1.06±.08	0.62±.14	8	0.997	0.113	17
1.22±.07	0.60±.11	16	0.994	0.137	18
1.20±.07	0.63±.11	8	0.998	0.095	19

(2) List all equations of the type: $\log 1/C = a\log P + bE_s + c$. Order them first according to increasing values of a and then according to increasing values of b. List the system and type of perturbation for each equation.

(3) List all equations having $\log P_o$ values in the range 4–5 for drugs acting on chloroplasts.

(4) List all drugs containing the function



having $\log 1/C$ values >5.0 and activity against virus.

(5) List all equations correlating the action of aromatic amines or phenols which contain a term in σ^-. Order these alphabetically by system.

(6) List the systems perturbed by all amides ($CONH_2$) having $\log P$ values of $2 \pm .5$.

The above queries are only a few of the scores which one can readily imagine to be of value in a truly comprehensive quantitative structure–activity system. In addition to structure–activity equations correlating biochemical systems it is important to include in the data bank as many examples as possible of substituent effects in homogeneous organic reac-

tions which can be used as standards of comparison for biochemical processes.

Although we have not completed all of the computer programming needed to answer all of the above questions, the dynamic part of the system in Figure 1 is more or less in operation. Some examples of the kind of questions which can be used to formulate a comparative pharmacodynamics can now be considered.

Tables II and III list equations which resulted from the present data bank from the following query: list all equations linear in log P where activity is defined as log $1/C$ having intercepts <1 and having at least five data points per set. Study of this print-out yielded the conclusion (*9*) that the equations of this type now in hand fell into two major classes: those of Table II with slopes of mean and standard deviation of $1.07 \pm .14$ and those of Table III with slopes of $0.74 \pm .09$. The equations were divided into the two sets solely on the basis of slope without regard to the kind of biological action involved. It would seem that all of the biological processes of Table II might be brought about by membrane perturbation. Comparison of the equations of Table II with the following average equation (*10*) for hemolysis lends credence to this hypothesis:

$$\log 1/C = 0.93 \pm .17 \log P - 0.09 \pm .23 \quad (18)$$

The slope of Equation 18 is the mean value for 15 different sets of drugs (ionic and nonionic) causing hemolysis. The intercept is the mean value from seven sets of *neutral* drugs causing hemolysis.

Table III. Log 1/C

Type of Compound	*Biological Activity*
ROH	Denaturation horse heart cytochrome
ROH	Denaturation whale myoglobin
ROH	Denaturation α-chymotrypsinogen
Miscellaneous	Precipitation sheep liver nucleoprotein at 40°C in 30 minutes
Miscellaneous	100% inhibition bovine muscle succinate oxidase
Miscellaneous	100% inhibition sheep liver succinate oxidase
Miscellaneous	Precipitation sheep liver nucleoprotein at 40°C in 15 minutes
Phenols	Growth inhibition *M. tuberculosis*
Miscellaneous	15–20% inhibition bovine muscle succinate oxidase
ROH	Growth inhibition *S. aureus*
Miscellaneous	15–20% inhibition sheep liver succinate oxidase
Miscellaneous	Inhibition fibrin swelling
Phenols	Conversion of cytochrome P-450 to P-420
Anilines	Conversion of cytochrome P-450 to P-420

The mean slope for the 17 examples of Table II is 1.07 ± .14. The equations have been ordered with respect to intercept with an overall difference in this parameter of about 1 log unit. Thus, narcosis of tadpoles requires about one-tenth lower concentration of drug than the 50% inhibition of indophenol oxidase (Equation 3). The equation most nearly resembling the model equation (Equation 18) is Equation 6 correlating the structure–activity relationship between the concentration of ROH necessary to produce a 5-mv change in the rest potential of the lobster axon. This close relationship between hemolysis and nerve membrane perturbation has been noted by others using different techniques (*11, 12*). The relationships of Table II show that different sets of molecules acting on very different systems can be compared quickly in numerical terms.

Such relationships can be useful in designing synthetic membranes having properties similar to natural systems. For example, Equation 4 correlates the change in resistance caused by alcohols on potassium ion permeability of black lipid membrane (BLM) prepared from the lipid of sheep erythrocytes. The rather large negative intercept of Equation 4 indicates that three times the concentration of isolipophilic alcohol is needed to change the resistance of the BLM as is needed to cause hemolysis. Although the two processes are quite different, the role of hydrophobic forces in each can be compared.

The intercept of the equations is open to more variation than the slope. The value of the intercept will depend on the sensitivity of the system and the degree of response demanded by the investigator (*e.g.*,

$= a \log P + b$

a	b	n	r	s	*Equation*
0.52±.10	−0.76±.04	5	0.995	0.026	20
0.84±.19	−0.58±.10	7	0.982	0.089	21
0.70±.14	−0.45±.07	8	0.982	0.071	22
0.82±.11	−0.26±.10	13	0.979	0.132	23
0.75±.15	−0.22±.16	14	0.951	0.205	24
0.76±.14	−0.19±.15	14	0.957	0.194	25
0.80±.13	−0.17±.12	14	0.970	0.132	26
0.78±.06	−0.09±.27	14	0.992	0.133	27
0.77±.12	0.05±.12	14	0.970	0.162	28
0.65±.12	0.06±.08	9	0.979	0.089	29
0.80±.14	0.10±.15	14	0.962	0.191	30
0.86±.32	0.21±.31	8	0.971	0.224	31
0.57±.08	0.36±.19	13	0.979	0.132	32
0.66±.15	0.36±.21	7	0.981	0.087	33

ED_{100}, ED_{50}, etc.). This effect can be seen by comparing Equations 5 and 6. Twice the concentration of isolipophilic compound is needed to produce a 10-mv change as is needed to produce a 5-mv change in the rest potential of the nerve. A similar effect is apparent in Equations 7 and 12. However, two different kinds of hearts are involved here, and the difference may in part be the result of the different types of tissue.

The slopes and intercepts of Equations 8 and 9 are quite close, indicating that the molecular probes inhibiting the two superficially different processes are probably operating in the same way at the molecular level. Inhibition of luminescence appears to be different from inhibition of bacterial growth (compare Equation 9 with Equations 27 and 29). In various published (*13*) and unpublished results linear relationships with slopes of about 0.7 for the inhibition of bacterial growth have been found.

Equation 13 correlates a mixed set of compounds which produce an abnormal type of mitosis resembling that caused by colchicine. The role of hydrophobic forces causing this kind of mitotic activity is closely related to that causing hemolysis (Equation 18), inhibiting quinea pig ileum (Equation 17), I_{50} red blood cell oxygen consumption, etc. The results of Table II indicate that Equation 18 can be used as a model for nonspecific membrane perturbation in various systems.

In Table III a different set of linear equations has been grouped. The mean slope for the 14 examples is $0.74 \pm .09$. The molecular probes used to obtain these equations indicate a significantly different dependence of biological response on log P. This mean slope is closer to that found for the binding of organic compounds by proteins (*14*). In fact, except for Equations 27 and 29, more or less isolated proteins constitute the biological systems of Table III. These systems are no doubt less lipophilic than the membranes of red blood cells or nerve cells which both contain 50% or more lipids. Of course, the highest concentrations (lowest intercepts) of isolipophilic molecules are required for the processes involving extensive denaturation or 100% enzymic inhibition (Equations 20–26).

Examples showing the relationship between the value of the intercept and the degree of response can be seen by comparing Equations 24 and 25 with 28 and 30. The differences in the two sets of intercepts are 0.27 and 0.29, indicating that a twofold increase in concentration is required to cause 100% inhibition of succinate oxidase over 15–20% inhibition.

Equations 32 and 33 have been placed on a log P basis and are therefore different from those published by Ichikawa and Yamano (*15*). In this study using rat liver microsome cytochrome P-450, there is little difference between the slopes or intercepts for the two different sets of congeners. In fact, a single equation could be used to correlate the 21

compounds. This indicates that hydrophobic forces alone (as operationally defined by log P) promote the conversion of P-450 → P-420.

Equations 27 and 29 are very much like Equations 32 and 33, suggesting that perturbation of the oxidative enzyme systems may be responsible for growth inhibition of the bacteria.

In summary, the equations of Tables II and III begin to sort out two general types of systems, each of which displays a different kind of nonspecific response to lipophilic molecular probes.

The larger the value of the intercept, the more sensitive is the system or the more specific the pharmacophoric function of the set of congeners. The following set of three equations was selected for the data bank to illustrate how more sensitive processes can be compared *via* extrathermodynamic correlations.

1-to-1 Binding of Miscellaneous Compounds by Bovine Hemoglobin (*16*)

$$\log 1/C = 0.71(\pm.13)\log P + 1.51(\pm.33) \qquad n = 17 \quad r = 0.950 \quad s = 0.160 \qquad (34)$$

I_{50} of Hydroxyindole-*O*-Methyltransferase by *N*-Acyltryptamine Amines (*17*)

$$\log 1/C = 0.60(\pm.10)\log P + 1.49(\pm.42) \qquad n = 21 \quad r = 0.948 \quad s = 0.170 \qquad (35)$$

75% Inhibition of Multiplication of Influenza B. Virus by Benzimidazoles

$$\log 1/C = 0.58(\pm.17)\log P + 1.58(\pm.46) \qquad n = 15 \quad r = 0.903 \quad s = 0.210 \qquad (36)$$

While the above three examples were picked because of their similarity in both slope and intercept, they are not necessarily identical processes at the molecular level. However, they must be quite similar. While it is known that in Equation 34 one molecule of substrate is bound to one molecule of protein, this is not known to be the case for Equations 35 and 36. There is a fair probability that this is true for Equation 35. The parameters in Equation 35 are slightly different from those derived by Lien, Hussain, and Tong (*17*) because of slight differences in log P values. The slopes of the above three equations are 0.6 ± .1. Such values are commonly observed for nonspecific binding of organic compounds by a variety of proteins and proteinaceous material (*3, 14*). Using Equations 34 and 35 as points of reference, some feeling about the inhibition of virus by benzimidazoles can be formed. Equation 36 has been formulated from the data in Table IV. Comparison of intercepts shows that this is not a very specific inhibitory process. It would be interesting to know how many benzimidazoles are bound per virus subunit so that a better comparison with Equation 34 could be made.

Table IV. 75% Inhibition Influenza B. Virus by Benzimidazoles

Compound	*log* P[a]	*log 1/C*		
		obsd.[b]	*calcd.*	$\lvert\Delta log\ 1/C\rvert$
1-Methyl	1.84	2.14[c]	2.65	0.51
Benzimidazole	1.34	2.46	2.36	0.10
2-Methyl	1.84	2.51	2.65	0.14
5-Methyl	1.84	2.72	2.65	0.07
5,6-Dimethyl	2.34	2.72	2.94	0.22
4,6-Dimethyl	2.34	2.82	2.94	0.12
2,5-Dimethyl	2.34	2.89	2.94	0.05
4,5-Dimethyl	2.34	2.96	2.94	0.02
2,4,6-Trimethyl	2.84	3.05	3.24	0.19
2,4,5-Trimethyl	2.84	3.20	3.24	0.04
5,6-Diethyl	3.34	3.39	3.53	0.14
2-Propyl-5-methyl	3.34	3.60	3.53	0.07
2,4,5,6,7-Pentamethyl	3.84	3.66	3.82	0.16
2-Ethyl,5-methyl	2.84	3.74	3.24	0.50
2-Butyl,5-methyl	3.84	3.77	3.82	0.05
2-Propyl,5-methyl	3.14	3.77	3.41	0.36

[a] Based on the value of 1.34 for benzimidazole.
[b] The value of 0.5 was added to this parent compound for each CH_3 or CH_2 to obtain other log P values.
[c] This point omitted since it is the only N-methyl derivative and it is a poor fit.

In perusing our present data bank for equations linear in log P with higher intercepts one finds that the hypnotic drugs make an interesting group for comparison. Equations 37 and 38 are derived from the data in Tables V and VI.

Minimum Hypnotic Dose in Mouse of N,N-Alkylarylureas (*19*)

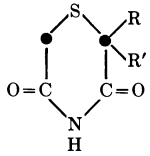
$$\log 1/C = 0.53(\pm.08)\log P + 2.44(\pm.10) \qquad n = 27,\ r = 0.945,\ s = 0.108 \qquad (37)$$

Minimum Hypnotic Dose in Rat of 5,5-Barbiturates (*21*)

$$\log 1/C = 0.57(\pm.21)\log P + 2.44(\pm.40) \qquad n = 15,\ r = 0.851,\ s = 0.156 \qquad (38)$$

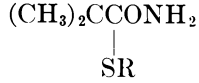
The great similarity of Equations 37 and 38 highlights the common mechanism of action of two superficially different types of amides. In each of the above examples, compounds having log P values below the optimum of log P_o were studied. Hence, addition of a term in $(\log P)^2$ did not result in an improved correlation in either case. The confidence intervals on the intercepts of Equations 37 and 38 are tighter than those on eight sets of parabolic equations correlating hypnotic activity of barbiturates in a variety of animals (*22*). However, for six of the parabolic equations with moderately good confidence intervals, a mean intercept

of 2.1 ± .3 is found. The agreement with Equations 37 and 38 is quite good, especially when one considers that the experimental conditions and the type of animals used were quite varied. The intercepts (*22*) (numbers under formulas) from parabolic correlation equations for the following three types of hypnotic amides are close to those of Equations



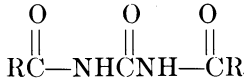
2.50(±.67)

II



1.98(±.54)

III



1.89(±.10)

IV

37 and 38. The above examples with hypnotics indicate that numerical comparisons can be made with quite different systems in a general way. In this way one gets a view of the forest without being too concerned about the individual trees.

Higher intercepts indicating higher sensitivity and specificity are contained in the following examples.

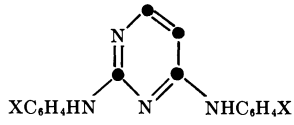
Killing of *M. fructicola* spores by Benzoquinones (*23*)

$$\log 1/C = 0.88(\pm.43)\log P + 3.53(\pm.80) \qquad n = 10,\ r = 0.859,\ s = 0.579 \qquad (39)$$

Inhibition of *S. typhosa* by Aromatic Amines (*13*)

$$\log 1/C = 0.58(\pm.12)\log P + 0.97(\pm.28) \qquad n = 15,\ r = 0.941,\ s = 0.136 \qquad (40)$$

I_{50} *C. albicans* (*23*) by



$$\log 1/C = 0.50(\pm.15)\log P + 4.15(\pm.35) \qquad n = 8,\ r = 0.957,\ s = 0.222 \qquad (41)$$

ED_{50} *S. aureus* in Mice of Phenoxypenicillins (*24*)

$$\log 1/C = -0.46(\pm.10)\log P\text{-ion} + 4.85(\pm.10) \quad \begin{matrix} n & r & s \\ 21 & 0.912 & 0.187 \end{matrix} \quad (42)$$

Even though Equation 39 is not a very sharp correlation (probably because it is lacking a suitable electronic term), it does serve to illustrate in isolipophilic terms the high intrinsic toxicity of quinones to fungi. Comparison of Equations 40 and 41 shows that simple aromatic amines do not have a high intrinsic toxicity. Thus the pyrimidine moiety, when

Table V. Hypnosis of Mice by *N,N*-Alkylarylureas[a]

			log 1/C		
	Compound	*log* P[b]	*obsd.*	*calcd.*	$\lvert\Delta log\ 1/C\rvert$
Me	Phenyl	0.42	2.46	2.67	0.21
Me	4-Anisyl	0.42	2.62	2.67	0.05
Me	2-Anisyl	0.42	2.67	2.67	0.00
Et	Phenyl	0.92	2.80	2.93	0.13
Me	2-Tolyl	0.92	2.82	2.93	0.11
Me	3-Anisyl	0.42	2.85	2.67	0.18
Me	4-EtO-phenyl	0.92	2.85	2.93	0.08
Et	4-Anisyl	0.92	2.87	2.93	0.06
Me	4-Tolyl	0.92	2.90	2.93	0.03
Et	2-Anisyl	0.92	2.93	2.93	0.00
Et	4-EtO-phenyl	1.42	3.00	3.20	0.20
Me	3-Tolyl	0.92	3.00	2.93	0.07
Me	4-EtO-phenyl	0.92	3.06	2.93	0.13
Me	2-EtO-phenyl	0.92	3.09	2.93	0.16
Et	3-Anisyl	0.92	3.15	2.93	0.22
Et	3-Tolyl	1.42	3.19	3.20	0.01
Et	2-EtO-phenyl	1.42	3.19	3.20	0.01
Et	4-Tolyl	1.42	3.19	3.20	0.01
Pr	Phenyl	1.42	3.19	3.20	0.01
Et	2-Tolyl	1.42	3.26	3.20	0.06
Et	3-EtO-phenyl	1.42	3.26	3.20	0.06
Pr	4-Tolyl	1.92	3.44	3.45	0.03
Bu	Phenyl	1.92	3.48	3.47	0.01
Pr	3-Tolyl	1.92	3.51	3.47	0.04
Pr	2-Tolyl	1.92	3.55	3.47	0.08
Bu	4-Tolyl	2.42	3.60	3.73	0.13
Bu	2-Tolyl	2.42	3.00	3.73	0.07

[a] From Ref. *19*.
[b] Log *P* values based on value of 0.42 for $C_6H_5N(CH_3)-CONH_2$. The OCH_3 group is given a value (*20*) of 0.00. Substituents in ortho, meta, and para positions are all given same π values. Adding a term in σ for benzene ring substitutuents did not improve the correlation nor did the addition of a term in $(\log P)^2$.

Table VI. Hypnosis of Rats by 5,5-Barbiturates[a]

			log 1/C		
	Compound	*log* P[b]	*obsd.*	*calcd.*	$\lvert\Delta log\ 1/C\rvert$
Et	Et	0.65	2.79	2.81	0.02
Et	Phenyl	1.42	3.12	3.25	0.13
Et	Amyl	2.15	3.45	3.67	0.22
Et	Isoamyl	1.95	3.50	3.56	0.06
Et	2-Me-butyl	1.95	3.45	3.56	0.11
Et	1-Me-butyl	1.95	3.81	3.56	0.25
Et	1-Et-propyl	1.95	3.81	3.56	0.25
Et	1,2-Di-me-propyl	1.75	3.45	3.44	0.00
Et	Cyclopentyl	1.79	3.45	3.45	0.00
Et	Butyl	1.65	3.33	3.39	0.06
Et	Isobutyl	1.45	3.28	3.27	0.01
Et	*sec*-Butyl	1.45	3.63[c]	3.27	0.36
Et	1-Me-amyl	2.45	3.60	3.84	0.24
Allyl	1-Me-butyl	2.15	3.83	3.67	0.16
Allyl	1-Et-propyl	2.15	3.77	3.67	0.10
Allyl	Cyclopentyl	1.99	3.67	3.58	0.09

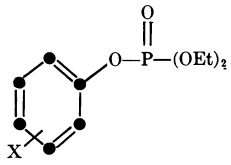
[a] From Ref. *21*.
[b] *See* Ref. *22* for experimental and calculated values.
[c] This data point not used in deriving Equation 38.

attached to the aromatic amino group, yields a function of high intrinsic activity. Comparisons of this type done early in drug modification studies should enable the drug designer to decide whether or not he is pursuing a false lead.

Equation 42 is included because it has a negative slope—*i.e.*, this set of congeners was selected so that all members had a superoptimal lipophilic character. Evidence now indicates (*22*, *25–27*) that the general relationship one should expect between log $1/C$ and log P is parabolic. The apex of the parabola has been termed log P_o. Molecules having this value of log P have ideal lipophilic character for the system under consideration. Since all of the above equations are linear in log P (the addition of a term in $(\log P)^2$ does not improve the correlation), greater activity could have been obtained in each example by designing molecules with better log P values. In all examples except Equation 42 this means increasing the lipophilic character. Equation 42 calls for less lipophilic molecules.

There are *very* few examples in our data bank where only an electronic term (σ, pK_a or M.O. parameters) suffices to correlate structure with activity. This accounts for the fact that up to this point molecular orbital calculations (*28*, *29*) of charge densities have yielded so few equations correlating chemical structure and biological activity in quantitative terms. However, the combination of electron distribution obtained *via*

quantum mechanical calculations and hydrophobic parameters does indicate the value of such calculations (*30, 31*). To illustrate the comparative value of the Hammett σ constant, consider the following equations correlating the structure–activity relationship of phosphate esters, V:



V

Hydrolysis (42 in Aqueous Solution, pH 7.6, 37°C)

	n	r	
$\log k = 1.96\ \sigma^- - 6.62$	4	0.966	(42)

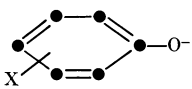
in vitro Inhibition Flyhead Cholinesterase

	n	r	
$-\log I_{50} = 2.49\ \sigma^- + 4.18$	8	0.985	(43)

in vivo LD_{50} House Flies

	n	r	
$-\log I_{50} = 2.42\ \sigma^- + 0.26\pi - 0.60$	8	0.987	(44)

In Equations 43 and 44, σ^- gives much better correlations than σ. For this reason it has also been used in Equation 42 although here it gives a slightly lower correlation, possibly because of the great difficulty in studying this reaction under such mild conditions and because only one function with a σ^- different from σ was studied. One would expect σ^- to be the parameter of choice since the leaving group in the reaction of the phosphate esters with cholinesterase is the phenoxide ion,



and functions delocalizing the negative charge *via* resonance should be more effective in increasing the hydrolytic rate. The work on which Equations 43 and 44 are based is from the pioneering studies of Metcalf and Fukuto (*32, 33*). The value of ρ in each of the above equations is consistent, indicating nucleophilic displacement as the common characteristic of the three reactions. It is noteworthy that ρ is somewhat higher in Equations 43 and 44. This is in line with the finding that ρ for a given reaction is higher in more apolar solvents (*34*). Equation 43 correlates

only para isomers while Equation 44 correlates both meta and para isomers. The meta isomers which are very poorly accommodated by Equation 43 can be correlated if a steric parameter is included for these functions (*35*) as follows:

$$-\log I_{50} = 0.55\, E_s^m + 2.45\, \sigma^- + 4.82 \qquad n = 13 \qquad r = 0.962 \qquad (45)$$

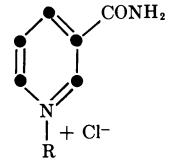
The negative coefficient with E_s^m (Taft steric parameter for the meta function) indicates that large groups increase the effectiveness of the inhibitor. The value for ρ is the same as in Equations 43 and 44. Equations 43 and 45 indicate that there is more specificity involved in the *in vitro* inhibition of cholinesterase than in the killing of flies. The closeness of the values for ρ in Equations 42 and 43 would indicate that a variety of nucleophiles might react with the phosphates with much the same dependence on σ^-. Therefore it seems likely that in the killing of flies described by Equation 44, the inhibition of a variety of less sterically demanding enzymes containing nucleophilic functions may be occurring and that one must not conclude that because ρ has the same value in Equations 43 and 44, killing is occurring by inhibition of cholinesterase alone. However, it is most encouraging for those interested in quantitative structure–activity work that the electronic effect of substituents on the reaction of phosphate esters with nucleophiles is much the same in the complex milieu of a fly as in isolated systems. Moreover, the read-out from the fly comes in the form of an LD_{50} which is not exactly comparable with a spectrophotometer reading. However, from the analysis of the perturbations of the fly, considerable information is obtained about the events at the molecular level in the fly.

As might be expected, in the fly work (Equation 44) activity depends slightly on the lipophilic character of the phosphate as indicated by need for a term in π. The small coefficient with this term indicates this dependence is not great.

Having a data bank which contains extrathermodynamic equations on both pure organic reactions as well as biochemical reactions is important in gaining insight into mechanism of action as Equation 42 illustrates. The type of σ constant best suited to correlate the data is important as well as the value and the sign of ρ. Examples in which σ^+ has proved to be a better parameter than σ are also known (*36*).

Turning now to steric factors, a print-out of all equations in the data bank containing a term in E_s and ordered on δ (the coefficient with E_s) yields a group of equations from which the following three examples were chosen because of the great similarity in the value of δ and the quite divergent character of the biochemical processes. Equations 47–48 were

Table VII. Inhibition of L-α-Glycerophosphate Dehydrogenase by



			log 1/C		
R	π	E_s	*obsd.*[a]	*calcd.*	|Δlog 1/C|
Methyl	0.50	0.00	1.82	1.76	0.06
Ethyl	1.00	−0.07	1.73	1.84	0.11
Propyl	1.50	−0.36	1.64	1.37	0.27
Butyl	2.00	−0.40	1.57	1.55	0.02
Pentyl	2.50	−0.40	1.82	1.79	0.03
Hexyl	3.00	−0.40	1.89	2.05	0.16
Heptyl	3.50	−0.40	2.15	2.31	0.16
Octyl	4.00	−0.40	2.38	2.57	0.19
Nonyl	4.50	−0.40	2.81	2.83	0.02
Decyl	5.00	−0.40	3.15	3.09	0.06
Undecyl	5.50	−0.40	3.38	3.35	0.03
Dodecyl	6.00	−0.40	3.78	3.61	0.17

[a] From Ref. *37*.

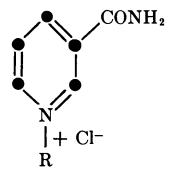
derived from the data in Tables VII–VIII. The parameters used in Equation 46 have been previously published (*5*).

Chymotrypsin Hydrolysis of $O_2NC_6H_4COOR$ (*5*)

$$\log k_2/K_m = 1.76(\pm.42)E_s + 0.79(\pm.39)\pi + 2.23(\pm.23) \qquad (46)$$

n	*r*	*s*
8	0.981	0.201

Inhibition (*37*) of L-α-Glycerophosphate Dehydrogenase by



$$\log 1/C = 2.53(\pm1.0)E_s + 0.52(\pm.08)\pi + 1.50(\pm.27) \qquad (47)$$

n	*r*	*s*
12	0.983	0.153

Cholinergic ED_{50} for $(X\text{–}CH_2COOCH_2CH_2N^+(CH_3)_3$ in Guinea Pig Ileum (*38*)

$$\log 1/C = 2.09(\pm 1.4)E_s + 1.74(\pm 1.7)\pi + 4.04(\pm .12) \quad (48)$$

$n = 6$, $r = 0.952$, $s = 0.246$

Except for Equation 47, each of the above examples contains fewer data points than one would like to see with equations containing three disposable parameters. However, note the similar values of δ. For two of the cases (Equations 46 and 48) this is not surprising since carbonyl functions are no doubt involved. The confidence interval on the E_s term in Equation 47 is rather wide so that one cannot be sure that the steric demands of this enzyme are the same as the receptors in the other systems.

While not much weight can be placed on the results in the above three systems because of the few data points involved, they do illustrate how comparisons can be made of quite different substrates being acted upon by very different systems. The fact that the values of δ in the above equations are not unreasonable can be seen by comparison with the values obtained from simple organic reactions. For the methanolysis, 1-propanolysis, and 2-propanolysis of $RCO_2C_{10}H_{16}$, the following respective δ values were found (39): 1.4, 1.7, 1.9. The reasonable agreement between Taft's δ values in biochemical processes and simple organic reactions under homogeneous conditions suggests that E_s may be a parameter of real importance in drug design (*6*).

What has the very limited experience of the past few years using substituent constants and regression analysis contributed to the problem of drug modification as outlined for derivatives of Figure 1? Possibly the most important fact is that the hydrophobic character of the members of almost any congeneric set of drugs must be considered even in

Table VIII. Cholinergic ED_{50} Guinea Pig Ileum[a] of $X\text{-}CH_2COOCH_2CH_2N^+(CH_3)_3$

			log 1/C		
X	π[b]	E_s[c]	*obsd.*	*calcd.*	$\lvert\Delta log\ 1/C\rvert$
H	0.00	1.24	6.70	6.63	0.07
CH_3	0.50	0.00	5.00	4.91	0.09
F	−0.17	0.78	5.22	5.37	0.15
Cl	0.39	0.27	5.19	5.28	0.09
Br	0.60	0.08	5.55	5.25	0.30
I	1.00	−0.16	5.23	5.45	0.22

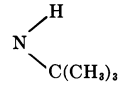
[a] From Ref. *38*.
[b] From Ref. *40*.
[c] From Ref. *41*.

in vitro studies with isolated tissue or enzymes. Moreover, there is abundant evidence that activity depends parabolically on hydrophobic character as defined by log P. One of the most powerful ways to eliminate many of the 2,500,000,000 analogs of I is to determine log P_o as soon as possible in the structure–activity study. The number of derivatives boils down quickly depending on how narrow the limits on this value can be set. If, for example, log $P_o = 4.0 \pm .3$, one can with some rational confidence neglect the synthesis and testing of derivatives whose calculated log P values fall outside the range 3.5–4.5. The additive natures (*3*) of log P and π make it possible to calculate log P values before synthesis. One way to begin to define log P_o would be to make 10 derivatives at a position well removed (say the 4- or 5-) from the side chain. This set could be selected so that variations in σ and E_s would also be present in the substituents. Fitting the 10 data points to Equation 50 should not only define log P_o but also give some information on steric and

$$\log 1/C = -k_1\pi^2 + k_2\pi + k_3\sigma + k_4E_s + k_5 \qquad (50)$$

electronic effects as well. Substituent effects on each of the other positions can be analyzed in the same way. With a well behaved system (many are not) one can begin to eliminate many analogs as one finds the signs and values of the coefficients with σ and E_s.

For example, assume that π_o, the optimum lipophilic character for substituents on the amide of 1-naphthoic acid, is found to be 2.0. Further assume for the hypothetical biochemical process that E_s has no role for substituents at the 3- and 4-positions and that ρ has been found to be -1.2. These facts greatly limit the choice of derivatives to be made. There is little point in making 3-derivatives since only the amino group in this position has a negative value and it is not very great. The dialkylamino function in the 4-position is the most reasonable choice of the above mentioned 50 functions. For $N(CH_3)_2$, $\pi = \sim 0.2$. Hence the optimum function would be something like $N(CH_2CH_2CH_3)_2$ with calculated π of 2.2. Now, of course, it is well known that such functions are easily dealkylated by microsomes, the hydrogens adjacent to N being most easily oxidized. One might try to circumvent this problem by studying alkylamino functions not having α-CH bonds—*e.g.*,



or one could use the less readily dealkylated —OR which has a lower σ_p value. Knowing π_o and taking advantage of the additivity of π greatly limits the possibilities.

If in the same problem ρ had turned out to have a positive value,

this would have greatly increased the possibilities since there are a variety of substituents with significant, positive σ_m values so that one could achieve more leverage in electronic terms by substituting in both the 3- and 4-positions at the same time. Assuming the same π_o value, one might try 4-NO_2, 3-CF_2CF_3. This combination would yield $\Sigma\sigma = 1.25$ and $\Sigma\pi = 2.0$. The possibilities for obtaining $\Sigma\pi = 2.0$ and a large $\Sigma\sigma$ would be manifold under these conditions. The final decision on which derivatives to make would depend on the cost of synthesis as well as the resistance of the various functions to unwanted metabolic transformations. For example, an SF_5 function would be most interesting because of its large σ and π values; however, the cost of synthesis might preclude its use.

Gaining maximum potency is only half the problem in drug design. The other half is to minimize unwanted side effects. Therefore, for the ultimate in maximization of the therapeutic index, regression analysis must also be applied to relate substituent effects to the various kinds of serious toxicity as these become evident.

To find the balance between maximum potency and minimum side effects from a potential group of billions of analogs which in a given test are being allowed to interact with a huge but unknown number of enzymes and critical membranes is indeed a monumental task. How many organic structures and their corresponding activities in the known biochemical systems can one mind contemplate? To the query, "Is the use of substituent constants and regression analysis worthwhile?", there is only one answer. We have been doing it and continue to do so. The only choice is between using a computer or one's solar plexus.

Acknowledgment

The author thanks W. R. Glave for making the correlation expressed in Equation 47 with data obtained from Ref. *37*.

Literature Cited

1. Smith, E. G., "The Wiswesser Line-Formula Chemical Notation," McGraw-Hill, New York, 1968.
2. Purcell, W. P., Singer, J. A., Sundarm, K., Parks, G. L., in "Medicinal Chemistry," A. Burger, Ed., 3rd ed., p. 164, Wiley-Interscience, New York, 1970.
3. Hansch, C., in "Drug Design," Ariëns, E. J., Ed., Vol. I, p. 271, Academic, New York, 1971.
4. Helmer, F., Kiehs, K., Hansch, C., *Biochemistry* (1968) **7**, 2858.
5. Hansch, C., Coats, E., *J. Pharm. Sci.* (1970) **59**, 731.
6. Kutter, E., Hansch, C., *Arch. Biochem. Biophys.* (1969) **135**, 126.
7. Hansch, C., Deutsch, E. W., *Biochim. Biophys. Acta* (1966) **112**, 381.
8. Hansch, C., *Ann. Repts. Med. Chem.* (1968) **1967**, 348.
9. Hansch, C., Dunn, W., *J. Pharm. Sci.* (1972) **61**, 1.

10. Hansch, C., Glave, W. R., *Mol. Pharmacol.* (1971) **7,** 337.
11. Seeman, P., *Inter. Rev. Neurobiol.* (1966) **9,** 145.
12. Schneider, H., *Biochim. Biophys. Acta* (1968) **163,** 451.
13. Lien, E. J., Hansch, C., Anderson, S. M., *J. Med. Chem.* (1968) **11,** 430.
14. Hansch, C., *Proc. Intern. Congr. Pharmacol.* (1969) **4,** 128.
15. Ichikawa, Y., Yamano, T., *Biochim. Biophys. Acta* (1967) **147,** 518.
16. Kiehs, K., Hansch, C., Moore, L., *Biochemistry* (1966) **5,** 2602.
17. Lien, E. J., Hussain, M., Tong, G. L., *J. Pharm. Sci.* (1970) **59,** 865.
18. Tamm, I., Folkers, K., Shunk, C. H., Heyl, D., Horsfall, F. L., Jr., *J. Exp. Med.* (1953) **98,** 245.
19. Hjort, A. M., de Beer, E. J., Buck, J. S., Johannes, S., Ide, W. S., *J. Pharmacol. Exp. Ther.* (1935) **55,** 152.
20. Fujita, T., Iwasa, J., Hansch, C., *J. Amer. Chem. Soc.* (1964) **86,** 5175.
21. Shonle, H. A., Keltch, A. K., Swanson, E. E., *J. Amer. Chem. Soc.* (1930) **52,** 2440.
22. Hansch, C., Steward, A. R., Anderson, S. M., Bentley, D., *J. Med. Chem.* (1968) **11,** 1.
23. Hansch, C., Lien, E. J., *J. Med. Chem.* (1971) **14,** 653.
24. Hansch, C., Steward, A. R., *J. Med. Chem.* (1964) **7,** 691.
25. Penniston, J. T., Beckett, L., Bentley, D. L., Hansch, C., *Mol. Pharmacol.* (1969) **5,** 333.
26. McFarland, J. W., *J. Med. Chem.* (1970) **13,** 1192.
27. Higuchi, T., Davis, S. S., *J. Pharm. Sci.* (1970) **59,** 1376.
28. Pullman, B., Pullman, A., "Quantum Biochemistry," Interscience, New York, 1963.
29. Kier, L. B., "Molecular Orbital Theory in Drug Research," Academic, New York, 1971.
30. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **86,** 1616.
31. Hansch, C., Deutsch, E. W., Smith, R. N., *J. Amer. Chem. Soc.* (1965) **87,** 2738.
32. Metcalf, R. L., Fukuto, T. R., *J. Econ. Entomol.* (1962) **55,** 340.
33. Fukuto, T. R., Metcalf, R. L., *J. Agr. Food Chem.* (1956) **4,** 930.
34. Jaffé, H. H., *Chem. Rev.* (1953) **53,** 191.
35. Hansch, C., *J. Org. Chem.* (1970) **35,** 620.
36. Hansch, C., *J. Med. Chem.* (1970) **13,** 964.
37. Kim, S. J., Anderson, B. M., *J. Biol. Chem.* (1968) **243,** 3351.
38. Chiou, C. Y., Sastry, B. V. R., *J. Pharmacol. Exp. Ther.* (1970) **172,** 351.
39. Taft, R. W., in "Steric Effects in Organic Chemistry," M. S. Newman, Ed., p. 644, Wiley, New York, 1956.
40. Hansch, C., Anderson, S. M., *J. Org. Chem.* (1967) **32,** 2583.
41. Kutter, E., Hansch, C., *J. Med. Chem.* (1969) **12,** 647.
42. Aldridge, W. N., Davison, A. N., *Biochem. J.* (1952) **51,** 62.

RECEIVED June 15, 1971.

Pitfalls in the Use of $\overline{\Pi}$ Constants

A. CANAS-RODRIGUEZ and M. S. TUTE

Research Division, Pfizer Ltd., Sandwich, Kent, England

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

The estimation of log P *by application of additivity principles presents some difficulties in heterocyclic and in sterically crowded or conformationally flexible molecules. The additivity rule can break down in such molecules not only because of unrecognized electronic effects but because of intramolecular hydrophobic bonding. We have compared experimentally determined log* P *values with the sum of* π *constants for diphenylmethane derivatives, including Benadryl, for anilines substituted on nitrogen with a flexible or rigid chain, for sterically crowded salicylic acid derivatives, and for both flexible and rigid structures in which two phenyl rings can interact through hydrophobic bonding. We conclude that measurement of log* P *is essential for such molecules and may give insight into their conformation in aqueous solution.*

The log P, or π, parameter is one of the most useful parameters of Hansch analysis because of the relevance of hydrophobic effects both to drug transport and to drug binding with some lipophilic site. Log P for any molecule more complex than a monosubstituted benzene is usually estimated by adding known π values for substituents to some molecular fragment (the nucleus) for which log P has been experimentally determined. Thus, in a molecule RXYZ the log P value may be estimated by adding log P_R for the nucleus, R, and π values π_x, π_y, π_z, for substituents assuming:

$$\log P = \log P_R + \pi_x + \pi_y + \pi_z$$

In applying this equation to polysubstituted benzenes, a correction for the dependence of π values on electronic interaction between substituents can usually be made by suitable choice of the fragment R and

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table I. Hydrophobicity Parameters in Hansch Equations

Transport Alone is Rate Determining	$\log BR = a \log P - b(\log P)^2 + \ldots + k$
Transport is not Rate Determining[a]	$\log BR = c\pi + \ldots + k$

[a] Where this applies, π refers to that part of the molecule desolvated on binding to the receptor.

relevant π values, so that an accurate estimate of log P, good enough for biological correlations, can be made.

We have been concerned with the applicability of this additivity rule to drug molecules, which are frequently heterocyclic compounds and which often are so flexible they can exist in more than one conformation in aqueous solution. In certain conformations, or in rigid but sterically crowded molecules, the groups chosen as X, Y, Z may interact with each other or with the nucleus R through space by intramolecular hydrophobic bonding. An experimental measurement of log P for such molecules can be compared with an additivity estimate to examine the relevance of the additivity equation and to give insight into intramolecular effects. With such data it may soon be possible to use log P to determine the preferred conformation of a drug molecule in aqueous solution.

Table I shows the various ways in which π and log P have been applied to Hansch analysis. In the first equation, log P refers to the complete molecule, and an optimum value is predicted from "random walk" theory when drug transport is rate determining. If this is applied as a model equation to complex molecules where additivity of π constants does not apply, log P must be measured, or large deviations will occur.

The second equation is often applicable to equilibrium situations such as those found *in vitro*. Sometimes, different dependencies on π are found for substituents in different parts of the molecule. The hydrolysis rates of substituted phenyl β-D-glucosides by emulsin, for instance, have been shown by Hansch (*1*) to depend on π for the para substituents but not on π for the meta substituents. Some of Baker's results (*2*) on the 5,6-disubstituted-2,4-diaminopyrimidines as inhibitors of dihydrofolate reductase can be correlated with π only by choosing π for the most lipophilic of the two substituents (*3*). Only a part of the molecule need be desolvated on combining with the receptor.

Figure 1 illustrates a common pitfall when log P is not measured. Apart from the fact that π and log P are not additive, one may be tempted to take π values from Hansch's work on phenoxyacetic acids, for example, and apply these without checking by log P measurement their validity for the series being studied. When π is taken from one series and applied

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

to another, an electrical component may be needed to compensate for the remoteness of the model. This may be especially necessary when the series has strongly polar or hydrogen-bonding functions. Wulfert's study (*4*) shows this effect clearly. The true π values for substituents on these quinazolones, obtained by log P measurement for each molecule, are related to π from the phenoxyacetic acid series and to the Hammett σ. Use of π from phenoxyacetic acids, without a check on log P, may result in the need for an electrical term, but one which cannot be interpreted. If transport is rate controlling, one might need π, π^2, σ, σ^2, and $\sigma\pi$ terms before the data will fit an equation. In practice there are usually not enough results available to accommodate so many terms without running out of degrees of freedom.

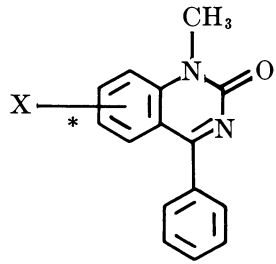


Figure 1. Dependence of π on σ (4)

$$\pi_{exp} = 1.09\ \pi_{POA} - 1.03\ \sigma + 0.10\ (n = 9,\ r = 0.97)$$

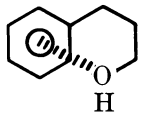
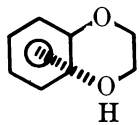
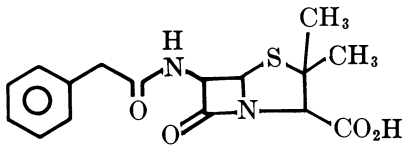
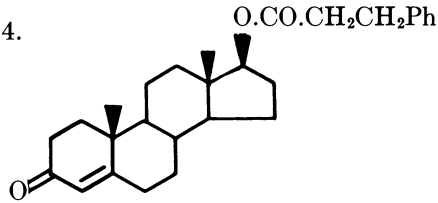
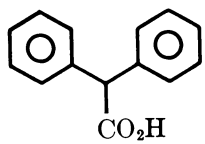
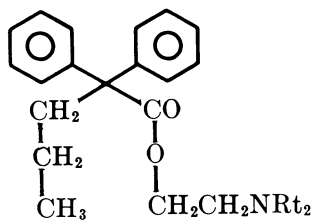
*where π_{POA} is π from phenoxyacetic acids (Hansch), σ is the Hammett constant, and * denotes para position*

Table II shows some literature examples of the breakdown in additivity which can occur in flexible molecules (Compounds 1–4) or in sterically crowded molecules such as the diphenylmethanes (Compounds 5 and 6). In the first four cases, a dipole is probably interacting with polarizable π-electrons of the aromatic ring. In Compound 5, the overall shape probably prevents water molecules from forming a solvate iceberg between the two rings—a situation which can be considered an intramolecular hydrophobic bond. In Compound 6 a combination of intramolecular hydrophobic effects and possibly interaction of the side chain dipole with one or both aromatic rings leads to a wide discrepancy between experimentally determined and calculated log P values.

Table III shows the structure of the antihistamine, Benadryl. Harms and Nauta (*5*), Kier (*6*), and Hansch (*7*) have suggested that the spatial relationship of the dialkylaminoalkyl side chain to the aromatic ring is an important determinant of activity in compounds of this type. We measured log P for Benadryl and found a value in close agreement with

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table II. Folded Molecules. $\Sigma\pi \neq \log P$

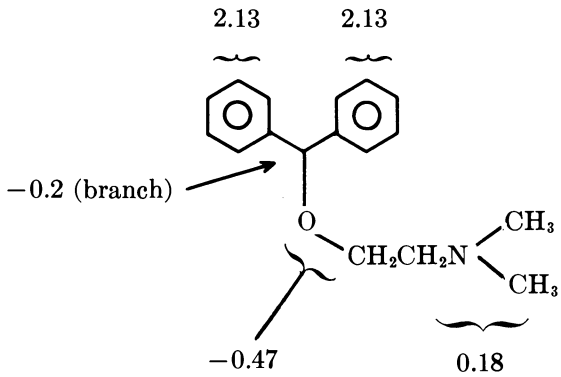
	Structure (Ref.)	*Log* P		$\Delta\pi$
		Expt.	*Additive*	
1.		1.88(*8*)	2.47	−0.59
2.		1.16(*9*)	1.45	−0.29
3.		1.70(*10*)[a] (PhCH$_2$−)	2.69 (PhCH$_2$−)	−0.99
4.		1.14(*11*)[a] (Ph)	2.13 (Ph)	−0.99
5.		2.06(*12*)	3.23	−1.17
6.		3.47(*9*)	6.27	−2.80

[a] Results extrapolated.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

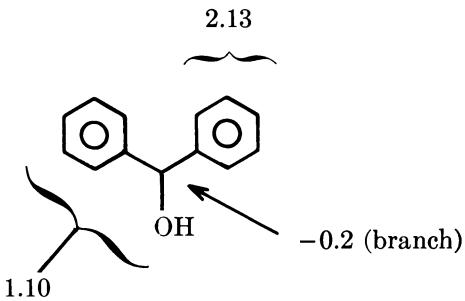
the previous measurements by Anderson (*13*). Calculation by summing π constants suggests at first that the side chain overlaps the aromatic nucleus, and the difference between measured and calculated log *P* values is similar to that found for flexible molecules such as phenylpropanol (Table II) or *N,N*-dimethylphenylpropylamine. The inference is that the polar group folds over the aromatic ring and that the dipole–π-electron interaction is further enhanced by an intramolecular hydrophobic effect in aqueous solution, resulting in a negative Δπ value. However,

Table III. Intramolecular Interaction in Benadryl



Σ π calculated = 3.77
log *P* measured = 3.30 (3.27, 3.40, Ref. *13*)
Δ π = −0.47 indicates chain folding?

but,



Σ π calculated = 3.03
log *P* measured = 2.03
Δ π = −1.00

Therefore, no chain folding can be presumed.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

in looking at the evidence a little more closely, is it fair, in summing π constants, to give a value of 2.13 to *each* phenyl ring when the two rings are so close that hydrophobic interaction between them could account for the breakdown in additivity?

We used diphenylmethanol to examine this point and found that in this compound also additivity does not hold. A definitive answer cannot therefore be given as to whether there is any hydrophobic driving force which causes Benadryl to exist in a folded conformation in aqueous solution. Moreover, with such strong interaction between the two aromatic rings, one cannot expect π values for any substituents in the ortho ring positions to bear any relation to π values in a simple benzene derivative.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

Table IV shows compounds which were of interest because of their anti-schistosomiasis effect. In all these compounds paramethylaniline is a common structural unit, and we believe that activity depends on transport to a site of metabolism, where hydroxylation of the methyl group occurs. An electronegative group ortho to the methyl is necessary for activity, as is an aminoalkyl side chain. In Compounds 1 and 3 this side chain is fixed in its position by ring formation; alternatively, the chain may be conformationally less rigid, as in Compounds 2 and 4. Log P measurements show that additivity principles do hold for all of these compounds and also suggest that an optimum log P exists; this log P_o is about 4.0, as in Compounds 1 and 2 which are the most active. Compound 3 is too lipophilic, and 4 is not lipophilic enough; both are less active than 1 and 2. It would seem that either a chloro or a nitro group would activate the methyl to hydroxylation; differences in activity between chloro and nitro derivatives are the result of different π effects of these groups on transport.

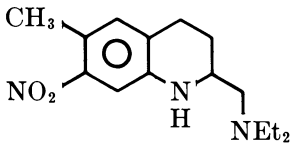
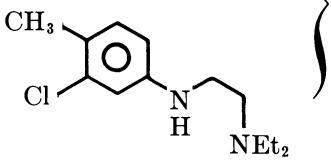
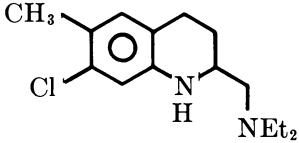
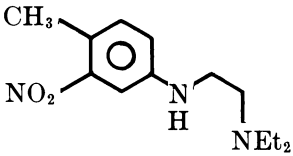
Table V shows some sterically crowded salicylic acid derivatives. Compound 1 is bibenzyl and is included for comparison. The $\Delta\pi$ value indicates some degree of interaction between the aromatic rings, which can be considered as an intramolecular hydrophobic effect. Log P is known for both salicylic acid and for phenyl methyl sulfide, and these values can be added to give the expected value for Compound 2. Experimentally, log P is smaller than this, so some intramolecular interaction probably occurs.

In Compound 3, a chlorine and a methyl group have been introduced. One would expect an increase of at least 1.0 log P unit since π (Me) is about 0.5, and π (Cl) is 0.6–1.0 according to electrical environment. In fact, the increase is only 0.4 log P unit.

Compound 4 is even more anomalous: here an isopropyl group is added ortho to the sulfide bridge, and again the methyl group. However, there is little difference in log P between this compound and Compound 2.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table IV. Schistosomiasis Compounds. $\Sigma\pi \simeq$ log *P*

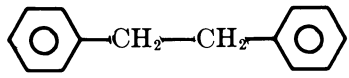
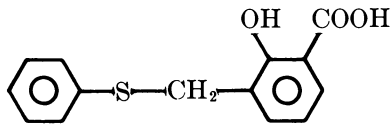
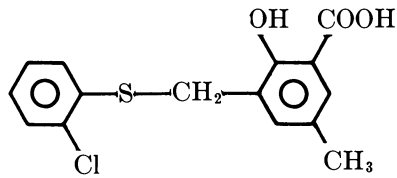
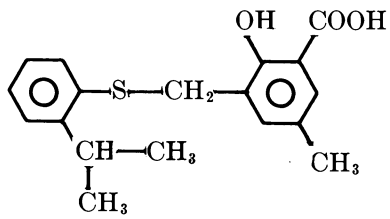
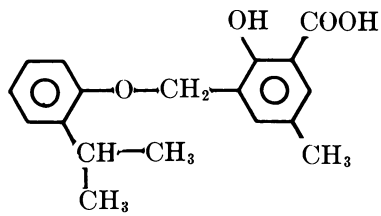
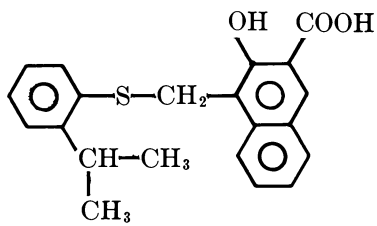
Structure	*Log* P	
	Expt.	*Additive*
1. 	4.19	4.42
2.  Optimum Lipophilicity	4.14	4.31
3.  Supraoptimal Lipophilicity		4.93
4.  Suboptimal Lipophilicity		3.63

Compound 5 is the oxygen analog of Compound 4 and has the greater log *P* value. Normally, sulfur compounds are considered to be more lipophilic than their oxygen analogs, so this result is the opposite of what one would expect on simple additivity grounds.

In Compound 6 an annelated ring is present. Compare this with Compound 4. Between these two, additivity seems to hold, the difference between π (Me) and π (annelation) being the same as the dif-

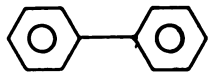
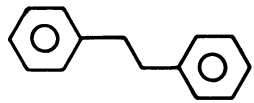
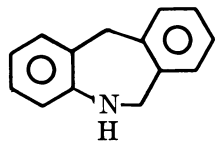
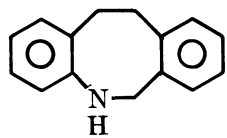
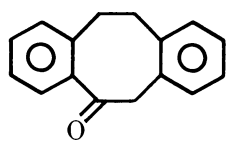
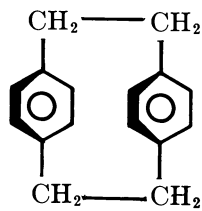
Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

Table V. Intramolecular Hydrophobic Bonds. $\Sigma\pi \neq \log P$

	Structure	*Log* P		$\Delta\pi$
		Expt.	*Additive*	
1.		4.79(*13*)	5.26	−0.47
2.		4.28	5.01	−0.73
3.		4.69	6.10	−1.41
4.		4.36	6.81	−2.45
5.		4.91	6.48	−1.57
6.		5.06	7.55	−2.49

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

Table VI. Mutual Shielding of Phenyl Rings. $\Sigma\pi \neq \log P$

	Structure	*Log* P		$\Delta\pi$
		Expt.	*Additive*	
1.		4.09(*9*)	4.26	−0.17
2.		4.79(*13*)	5.26	−0.47
3.		2.90	4.11	−1.21
4.		2.82	4.52	−1.70
5.		2.58	4.44	−1.86
6.		2.33	5.90	−3.57

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

ference in experimental log P between the two compounds. Can we presume that neither the methyl group nor the annelated ring is involved in any intramolecular effect? In crowded molecules such as these, log P/π additivity certainly cannot be expected to hold, but it is possible that measurement of log P can give insight into the conformation of the molecules in aqueous solution, which may be useful when considering exactly how the molecule interacts with the receptor.

Table VI shows a progression in the intramolecular hydrophobic interaction between two rings. In diphenyl, additivity holds. In diphenylethane, which has some conformational freedom, there is a small but quite definite interaction. In the tricyclic Compound 3, the rings are held at a definite angle to one another, with no possibility for rotation. Adding log P values for N-methylaniline and toluene, and making a correction for cyclic methylene groups, we obtain an estimate far greater than the experimental log P value.

In Compounds 4 and 5, the planes of the two rings approach one another even more closely, the dihedral angle between the planes is smaller, and additivity breaks down still more.

The last example is paracyclophane. This molecule has been studied well, and the two rings lie face to face so that considerable interaction is possible between the delocalized electrons. The dihedral angle is zero, and additivity is nonsensical.

We hope these examples will caution all who use π values as parameters in a multiple regression analysis. Whenever intramolecular interactions are possible, either because of steric crowding or conformational flexibility, it is essential to measure log P.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch003

Literature Cited

1. Hansch, C., Deutsch, E. W., Smith, R. N., *J. Amer. Chem. Soc.* (1965) **87**, 2738.
2. Baker, B. R., *J. Pharm. Sci.* (1964) **53**, 347.
3. Miller, E., Hansch, C., *J. Pharm. Sci.* (1967) **56**, 92.
4. Wulfert, E., Bolla, P., Mathieu, J., *Chim. Therap.* (1969) **4**, 246.
5. Harms, A. F., Nauta, W. T., *J. Med. Chem.* (1960) **2**, 57.
6. Kier, L. B., *J. Med. Chem.* (1968) **11**, 441.
7. Kutter, E., Hansch, C., *J. Med. Chem.* (1969) **12**, 647.
8. Fujita, T., Iwasa, J., Hansch, C., *J. Med. Chem.* (1965) **8**, 150.
9. Church, C., unpublished analysis.
10. Biagi, G. L., Barbaro, A. M., Gamba, M. F., Guerra, M. C., *J. Chromatog.* (1969) **41**, 371.
11. Biagi, G. L., Guerra, M. C., Barbaro, A. M., *J. Med. Chem.* (1970) **13**, 944.
12. Bowden, K., Young, R. C., *J. Med. Chem.* (1970) **13**, 225.
13. Anderson, S., unpublished analysis.

RECEIVED May 10, 1971.

4

Relationships between Partitioning Solvent Systems

ALBERT J. LEO

Pomona College, Claremont, Calif. 91711

Among the forces which can act between a small molecule and a macromolecule to elicit a biological response, none is more ubiquitous than hydrophobic bonding. Hydrophobic bonding is difficult to define, but it can be measured by using partition coefficient data. How a solute is distributed between water and any immiscible solvent is influenced by many foces besides hydrophobic bonding, but these polar effects are minimized using an alcohol/water system because they operate to some extent in both phases. From the regression equations relating 20 other systems to octanol/water, it is often possible to calculate the octanol P *value for a solute measured in another system. The hydrophobic bonding parameter, π, is generally additive, and reasonably accurate log* P *values can be estimated from molecular structure.*

Since partitioning between immiscible solvents is an equilibrium process, it should be possible to calculate partition coefficients for any solute between any two given solvents if we had a measure of the solvation forces involved. At present these forces are not that well characterized, and the reverse approach can be more enlightening—*i.e.*, a study of how the partition coefficient varies between the systems can yield valuable insight into the types and relative magnitude of the forces involved.

These forces of solvation, which result from ionic interaction, hydrogen bonds, hydrophobic "bonds," and other dipole–dipole attactions are important in biological studies not only because they dictate the rate at which any solute can pass through various lipid membranes but because they also influence binding to proteins, to serum albumin in transport throughout the body, and to enzyme surfaces (*1*).

Table I. Comparison

	Lipoid Phase	*Moles H_2O* ($\times 10^3$)
Increasing Lipophilicity ↓	1. *n*-Butyl alcohol	9440
	2. Cyclohexanol	6510
	3. 2-Butanone	5460
	4. 2-Pentanol	5320
	5. Primary pentanols[a]	5000
	6. Cyclohexanone	4490
	7. Octanol	2300
	8. Ethyl acetate	1620
	9. Methyl isobutyl ketone	950
	10. Oleyl alcohol	712

[a] Average of *n*- and iso-amyl alcohols.

The partition coefficient, P, is the ratio of the concentration of the solute in the lipoid phase divided by the concentration in the aqueous (polar) phase. If the solute species is the same in both phases, the solute activity in each phase must be equal (at equilibrium), and in dilute solution where the activity coefficient is close to unity the ratio of concentrations in the two phases must remain constant regardless of initial solute concentration.

Over 20 years ago Collander (*2*) showed that there was a linear relationship between the log P values measured in any two alcohol/water systems.

$$\log P_2 = a \log P_1 + b \tag{1}$$

This is not surprising when we consider that, in transferring from the polar aqueous phase to the lipoid phase, a solute molecule would in every case be exchanging "hydration" forces for those forces provided by an alkyl chain, plus a hydroxyl group. When the number of solvent systems is enlarged to include esters, ketones, some halogenated hydrocarbons, and aromatic and aliphatic hydrocarbons, no such simple relationship as expressed in Equation 1 holds, unless one restricts it to a limited class of solutes.

Since there are at least 20 solvent systems in common use in partitioning work and since any one of them can be compared with any other in an equation such as (1), it is important to choose one system as a standard to develop a uniform frame of reference. We have chosen the octanol/water system as P_1 for a number of reasons, one of the more important being that the greatest number of biologically interesting solute types have been measured in it.

of Solvent Lipophilicity

	Lipoid Phase	*Moles H_2O* ($\times$ *10^3*)
Increasing Lipophilicity ↓	11. Ether	690
	12. Isopentyl acetate	456
	13. Nitrobenzene	180
	14. Oils	72.5
	15. $CHCl_3$	68.4
	16. Benzene	26.0
	17. Toluene	25.6
	18. Xylene	18.8
	19. CCl_4	10.0
	20. Heptane	3.3
	21. Cyclohexane	2.5

The next important step is to establish some sort of lipophilic order in the solvent systems. At one time or another, various parameters have been proposed for such a scale—dipole moment, dielectric constant, solubility parameter, among others. For this particular application, we found that the solvent's lipophilic character could be measured by its inability to accommodate water molecules—*i.e.*, lipophilicity of a solvent can be measured by the reciprocal of the concentration in moles/liter of dissolved water at saturation.

Table I shows that when the solvent systems are ordered in this way, the primary butanols are the least lipophilic and oleyl alcohol is the most lipophilic of the solvents in the left column. Up to this point, the partition coefficients of various solutes are accommodated by Equation 1. The separation of the solvents in the right-hand column as more lipophilic is somewhat arbitrary, of course, but the partition coefficients measured in these systems are poorly correlated with any of the systems in the left column and generally cannot be correlated well with each other.

The free energy required to transfer a solute between the two phases can be factored conveniently into two components (*3*):

$$\log P = \frac{m\Delta\mu_L^\theta + j\Delta\mu_H^\theta}{2.3\,RT} \qquad (2)$$

In this extrathermodynamic expression, the chemical potential (per mole) is given the symbol, μ, and the solute is pictured as being divided into m lipophilic and j hydrophilic segments. Such a solute is hexamethylene glycol, where $m = 6$ and $j = 2$. Now one can predict the requirements needed for Equation 1 to hold.

Table II. Hydrophilic Solvent Systems Compared with Octanol[a]

Solvent System	*Slope a*	*Intercept b*	*n*	*r*	*s*
Primary butyl alcohols	0.697	+ 0.381	57	0.993	0.123
Cyclohexanol	0.745	+ 0.866	12	0.985	0.100
2-Butanone	0.493	+ 0.315	9	0.987	0.093
sec- & *tert*-Pentyl alcohols	0.982	+ 0.288	11	0.996	0.091
Primary pentyl alcohols	0.808	+ 0.271	19	0.987	0.161
Cyclohexanone	1.035	+ 0.896	10	0.972	0.340
Octanol	1.000	+ 0.000	std. of reference		
Ethyl acetate	0.932	+ 0.052	9	0.969	0.202
Methyl isobutyl ketone	1.094	+ 0.050	17	0.993	0.184
Oleyl alcohol	0.999	− 0.575	37	0.985	0.225

[a] *P* values for all of the solutes used to develop the equations are listed in *J. Org. Chem.* (1971) **36,** 1539; Microfilm edition

A linear relationship between log *P*'s will exist if one of two requirements is met: the primary solvation forces in the two solvent systems can be so similar that a variety of lipophilic and hydrophilic solute groups are accommodated proportionately, or the structural differences in the solute set being considered are such that one of the right hand terms in Equation 2 is essentially constant. The latter condition often applies to a homologous series where the hydrophilic group (an OH or CO_2H) contributes a constant component to the total transfer free energy.

Examples of how the first condition can be met are given in Table II, which lists the slopes and intercept values of each equation of type (1) relating the least lipophilic solvent systems to octanol/water. The number of solute *P* values used to establish each equation is listed under

Table III. Lipophilic Solvent Systems

	H-Donor Solutes				
Solvent System	*Slope*	*Intercept*	*n*	*r*	*s*
Ether	1.130	− 0.170	71	0.988	0.186
Isopentyl acetate	1.027	+ 0.072	22	0.986	0.209
Nitrobenzene	1.176	− 1.072	9	0.977	0.217
Oils	1.099	− 1.310	65	0.981	0.271
$CHCl_3$	1.126	− 1.343	28	0.967	0.308
Benzene	1.015	− 1.402	33	0.962	0.234
Toluene	1.135	− 1.777	22	0.980	0.194
Xylene	0.942	− 1.694	19	0.963	0.225
CCl_4	1.168	− 2.163	24	0.974	0.282
Heptane	1.056	− 2.851	10	0.764	0.916
Cyclohexane	0.675	− 1.842	26	0.761	0.503

[a] *P* values for all of the solutes used to develop the equations are listed in *J. Org. Chem.* (1971) **36,** 1539; Microfilm edition.

n; the correlation coefficient is *r;* the standard deviation is *s.* The solvation forces operating with various solutes must be similar since they all show a close relationship to octanol/water. This may seem surprising at first for these solvents are alcohols, esters, and ketones whose hydrogen bonding capabilities ought to vary considerably. However, so much hydrogen bonding capability is contributed by the water, which is present at about 1*M* level in even the most lipophilic solvent of the group, that it provides the equalizing effect on the solvation properties.

The second condition which can permit a linear relationship between log *P* values is illustrated in the equations in Table III which relate the more lipophilic systems to octanol/water. For the isopentyl acetate and the nitrobenzene systems, the correlation with the octanol system is quite good with a single equation, but the only solute values available were those from a single homologous series (the carboxylic acids), and we predict that a wider selection of solutes would result in a poor correlation.

When the log *P*'s of the rest of the lipophilic solvent systems are compared with octanol/water, a poor linear correlation results. The *r* values are generally below 0.8, and the standard deviations are generally in the range of 0.6 to 0.8 log units. In the individual solute deviations nearly all of the minus deviants consist of strong hydrogen bond donors while the plus deviants generally are hydrogen bond acceptors. In fact, if the solutes are first segregated into two groups, based on known donor or acceptor capability (*4*), two vastly improved correlations result. These appear in Table III.

No doubt it is an oversimplification to assign to hydrogen bonding forces all the transfer free energy of the hydrophilic groups as expressed in Equation 2. Other ion–dipole and dipole–dipole forces can influence

Compared with Octanol[a]

H-Acceptor Solutes				
Slope	*Intercept*	*n*	*r*	*s*
1.142	− 1.070	32	0.957	0.326
1.119	− 0.325	14	0.988	0.233
1.276	+ 0.171	21	0.976	0.251
1.223	− 0.573	19	0.958	0.291
1.398	− 0.922	14	0.971	0.274
1.027	− 0.595	21	0.986	0.230
1.207	− 0.219	11	0.959	0.347
1.848	− 2.223	11	0.954	0.534
1.063	− 0.734	30	0.957	0.360

the position of equilibrium, but whenever both types appeàr (*e.g.*, in a solute with both a cyano and a hydroxyl group), hydrogen bonding appears to exert the greater influence. This statement might not be valid if the polar phase were a solvent other than water.

Table IV. General Solute Classes

Group A: H-Donors		1. Carboxylic Acids
		2. Phenols
	"Neutral" in $CHCl_3$	3. Barbiturates
		4. Alcohols
		5. Amides (Neg. Subs., Not Di-*N*-Subs.)
		6. Sulfonamides
		7. Nitriles
		8. Imides
		9. Amides
Group B: H-Acceptors		10. Aromatic Amines (Not Di-*N*-Subs.)
		11. Miscellaneous Acceptors
		12. Aromatic Hydrocarbons
		13. Intramolecular H-Bonded Compounds[a]
		14. Ethers
		15. Esters
		16. Ketones
		17. Aliphatic Amines and Imines
		18. Tertiary Amines (Including Ring *N* Compounds)

[a] *e.g.*, *o*-nitrophenol.

From the present data the solute classification scheme shown in Table IV is proposed. The strongest hydrogen bond donors are listed at the top and the strongest acceptors at the bottom. At either end of the scale the order agrees with other methods of measurement (*5*, *6*). The order within the middle group (No. 3–9) is not as clear cut. In most of the lipophilic solvent systems studied, only two solute classes were required to obtain equations with good r values, and groups No. 3–9 are placed in the donor category. With $CHCl_3$ and CCl_4, a further subdivision seemed advisable, and these solute types were assigned a neutral classification.

Whenever the solute contains multiple functional groups, a donor usually predominates over an acceptor. Whenever donor–acceptor requirements can be met internally, the solute usually falls into class No. 13. Other influences are also discernable—*e.g.*, aliphatic amides are weak donors and are nearly as well predicted when grouped in the acceptor equation (in fact with the ether and oil systems, this is the preferred class). However, when amides are negatively substituted, such as trichloroacetamide, they must be correlated using the donor equation.

From the slopes and intercepts of the equations given in Tables II and III we can put some of the solvation forces on a more quantitative

basis. In Table II the slope is a measure of the solvents systems' sensitivity to changes in the lipophilic–hydrophilic balance of the solute. Butanol/water, as expected, has a low slope and low sensitivity. When this pair is saturated with one another, they are about as alike as two separate phases can be. Since log P measures the difference in transfer energy between the two phases, any change in solute character will register as only a small difference when compared with the same change if it is measured in the octanol/water system.

Increasing the hydrocarbon chain length in the solvent alcohol increases the dissimilarity of the two phases, and the sensitivity (*i.e.*, the slope value) increases. Apparently a maximum sensitivity is reached at octanol since the slope in the equation for oleyl alcohol is also 1.0.

If hydrogen bonding could be accounted for separately, all of the slopes in the equations comparing the lipophilic solvent systems to octanol would be close to 1.0. For example, cyclohexane values correlate poorly with octanol values when all solutes are taken together. Even for phenols within the H-donor group, the correlation coefficient is only 0.76. However, if we add a log term which measures the hydrogen bonding capability of phenols (*5*), the correlation coefficient increases to 0.98 and the slope becomes 1.0.

The transfer free energy (as measured by the slope value) for any solute in an alcohol/water system should decrease as the chain length of the solvent alcohol is shortened. Of course the same effect ought to be observed if we hold the oil phase constant and make the aqueous phase less hydrophilic. The equations in Table V show that for a limited set of barbiturates partitioned between diethyl ether and a 50–50 mixture of water and dimethylformamide, the slope, compared with the octanol standard, is only 0.4 whereas the ether/water system gives a slope for the donor solute group of 1.13. Thus, a 50% reduction in the protic character of the polar phase reduces the sensitivity of the system by a factor of 2.8.

Table V. Transfer Free Energy Reduced by Decreasing Protic Character of Aqueous Phase

System	*Correlation with Octanol/Water*
1. Diethyl ether/water	$\log P = 1.13 \log P_{(oct)} - 0.17$ $n = 71, r = .988, s = .186$
2. Ether/H_2O–DMF(*7*)	$\log P = 0.40 \log P_{(oct)} - 0.36$ $n = 6, r = .988, s = .058$

Reduction in Sensitivity to Solute Changes:

$$\frac{1.13}{0.40} = 2.8 \text{ fold.}$$

The data in Table VI show how the intercept value for each of the donor equations can also be used as a measure of the lipophilicity of the solvents. The intercept in the equation for any solvent system is the log P in that system for any solutes which are distributed equally between water and octanol—*i.e.*, solutes where log $P_{(oct)} = 0$. Therefore, a negative intercept means the solvent is more lipophilic than octanol, and a positive intercept means it is less lipophilic. This is more apparent when considering a homologous series—*e.g.*, the carboxylic acids as shown in Table VI.

The octanol log P values begin at -0.54 for formic, rise to -0.17 for acetic, and 0.33 for propionic. That is, it takes between two and three lipophilic methylene groups to balance the hydrophilic carboxyl group and allow octanol to share the solute equally with water. In oleyl alcohol/water it takes one additional methylene before a carboxylic acid is lipophilic enough to be equally shared—*i.e.*, log $P_{(oleylalc.)} = 0$ between propionic and butyric acids. In nitrobenzene/water it takes two additional methylenes, in benzene/water it takes three, and in carbon tetrachloride it takes about 4.5 additional methylenes before log $P = 0$.

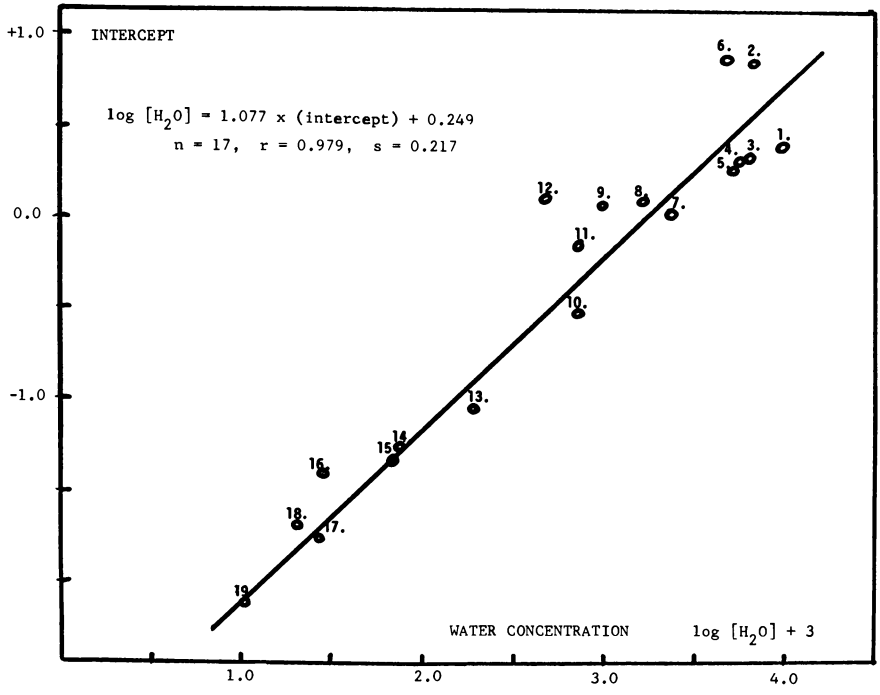


Figure 1. Water concentration vs. *intercepts from Equation 1. Numbers refer to solvent systems listed in Table I.*

Table VI. Homologous Carboxylic Acids

	Solvent System Lipophilicity →				
Log P	*Octanol*	*Oleyl Alcohol*	*Nitro-benzene*	*Benzene*	CCl_4
+1.0	Butyric				
+0.5	Propionic	Butyric			
0.0	Acetic	Propionic	Valeric	Valeric	
−0.5	Formic		Butyric	Butyric	Valeric
−1.0			Propionic	Propionic	Butyric

Earlier we arranged the solvents according to their water content at saturation and proposed that this constituted a lipophilicity scale. In effect, this equated lipophilicity with hydrophobicity. With the intercepts of the equations relating each system to octanol, we have a second method of measuring lipophilicity, and it is interesting to see how well they agree. In Figure 1 the intercepts (on a log scale) are plotted against the log [H_2O] and except for solvents No. 6 and No. 12 (cyclohexanone and iso-pentyl acetate) the agreement is satisfactory.

Conclusions

The octanol/water partition coefficient of a particular solute is not purely and simply a measure of the hydrophobic forces which that particular structure is capable of supplying to an aqueous environment. Hydrogen bonding and other dipole forces also contribute to the energy required to transfer the solute from water to octanol; however, since in this system both phases have an –OH group, there is not such a premium placed on the solute's H-bonding capability as there would be in a hydrocarbon/water system. When it is important to estimate the probable ratio of hydrophobic to polar forces in the interaction of the solute in question with various macromolecules or membrane surfaces, no single partitioning system can serve as a perfect model, and one must examine the variation of the values as measured in several systems. Deviations from the ideal equation, $\log P_2 = a \log P_1 + b$, can often be interpreted on the basis of the relative importance of polar solvation forces in any particular instance.

Literature Cited

1. Hansch, C., *Accounts Chem. Res.* (1969) **2,** 232.
2. Collander, R., *Acta Chem. Scand.* (1951) **5,** 774.
3. Cratin, P., *Ind. Eng. Chem.* (1968) **60,** 14.
4. Pimentel, G., McClellan, A., "The Hydrogen Bond," p. 229, Reinhold, New York, 1960.
5. Higuchi, T., Richards, J., Davies, S., Kamada, A., Hou, J., Nakano, M., Nakano, N., Pitman, I., *J. Pharm. Sci.* (1969) **58,** 661.
6. Taft, R., Gurka, D., Joris, L., van R. Schleyer, P., Rakshys, J., *J. Amer. Chem. Soc.* (1969) **91,** 4794, 4801.
7. Feltkamp, H., *Arzneim-Forsch.* (1965) **15,** 238.

RECEIVED June 15, 1971. Work supported under National Institutes of Health grant CA 11110 and contract No. 70-5115.

5

Partition Data of Chemotherapeutic and Steroid Agents Determined by Reversed-Phase Thin Layer Chromatography

G. L. BIAGI, A. M. BARBARO, and M. C. GUERRA

Istituto di Farmacologia e Farmacognosia dell'Università di Bologna, Italy

The chromatographic R_m value, related to the logarithm of the partition coefficient between the polar and the nonpolar phase of a chromatographic system, can be used to estimate the lipophilic character of compounds. R_m values of penicillins, cephalosporins, and testosterone esters are related linearly to the acetone concentration in the mobile phase. To obtain the R_m for each compound in a standard system, the theoretical R_m values at a given acetone concentration in the mobile phase were calculated by interpolation or extrapolation from the linear part of the curve. The good correlations between R_m and Hansch π values support Collander's findings; he showed a linear relationship between the partition coefficients obtained with two different sets of solvents. The nature of the phases involved in determining the partition data should not affect the correlation with biological activity.

The lipophilic nature of molecules seems to play a major role in determining the interaction of drugs with biological systems (*1*). Hansch and Fujita (*2*) developed a rational quantitative approach to structure–activity correlations by means of a substituent constant π. This parameter was defined as the difference in the logarithm of the octanol–water partition coefficients P of the substituted and unsubstituted compounds. For a substituent group, π is a constant, provided that introduction of the substituent into a molecule does not cause group interactions that would affect the partition coefficient of the molecule itself (*3*). Therefore, because of possible group interaction, calculated π values cannot be completely substituted for the experimentally determined octanol/water

partition coefficient P (*2*). To avoid the practical difficulties in determining the partition coefficient directly, Boyce and Milborrow (*4*) suggested the use of the chromatographic R_m value [$R_m = \log (1/R_f - 1)$]. This is related to the logarithm of the partition coefficient between the polar and the nonpolar phase of a chromatographic system (*5, 6, 7*). On the other hand, if there are no group interactions (*8*), the ΔR_m value, defined as the difference between the R_m values of the substituted and unsubstituted compounds, can be a parameter for estimating the lipophilicity of compounds.

In our laboratory R_m values were used as an expression of the lipophilic character of the molecules. Details of the chromatographic procedure have been described (*9, 10*). The compounds were partitioned between a nonpolar stationary phase and a polar mobile phase. The stationary nonpolar phases was obtained by impregnating a silica gel G layer with silicone DC 200 (350 cS) from Applied Science Laboratories.

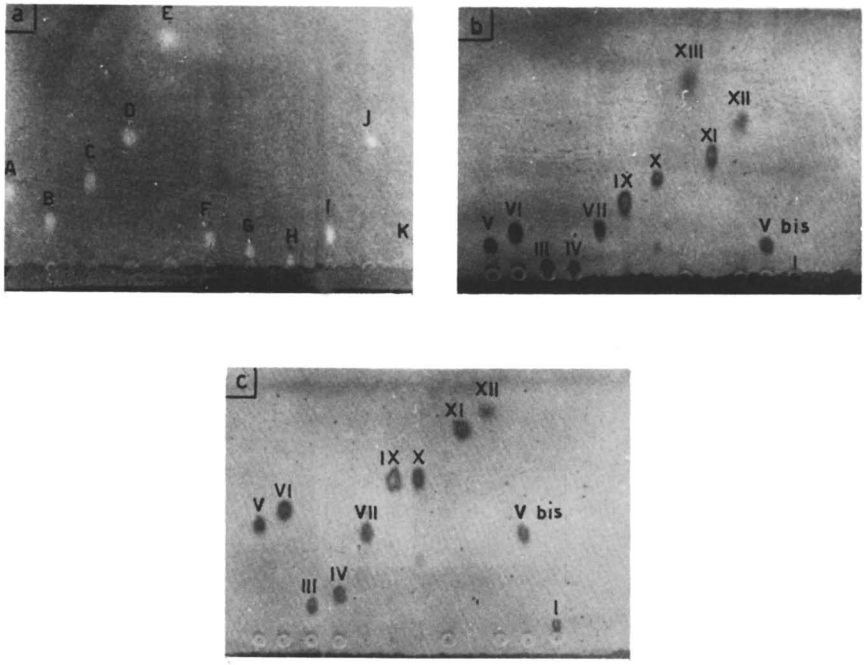


Figure 1. Reversed phase TLC of penicillins (a) and cephalosporins (b and c). In (a) the mobile phase is represented by 2% acetone–buffer: (A) Benzylpenicillin; (B) Phenoxymethylpenicillin; (C) Methicillin; (D) Ampicillin; (E) Carbenecillin; (F) Oxacillin; (G) Cloxacillin; (H) Dicloxacillin; (I) Phenethicillin; (J) Metampicillin; (K) Nafcillin. In (b) and (c) the mobile phase is represented by 10% buffer and acetone–buffer. The compounds indicated are listed in Table III. Compound V (bis) corresponds to acid cephaloram. In (c) Compound XIII is no longer detectable since it migrated with the solvent front.

The mobile polar phase was an aqueous buffer (sodium acetate–Veronal buffer 1/7*M* at pH 7.4) alone or mixed with various quantities of acetone. The compounds were dissolved in water, acetone, or ethanol (1–3 mg/ml) and spotted in 1 μliter amounts. The spots were detected by an alkaline solution of potassium permanganate. The penicillins could be detected also by iodine azide solution.

Results and Discussion

Chemotherapeutics. When the developed plates were sprayed with potassium permanganate, round spots appeared at various distances from the starting line. An example of reversed-phase thin layer chromatography (TLC) of penicillins is shown in Figure 1a (*10*). The most hydrophilic compound was carbenecillin, characterized by a longer migration with the polar mobile phase (acetone–buffer 2%). The most lipophilic compound appeared to be dicloxacillin, which essentially did not move from the starting line. An R_f value of 0 or 1 would indicate a compound most soluble in the nonpolar stationary phase or in the polar mobile phase. R_f values of the penicillins were plotted in Figure 2 against the composition of the mobile phase. These values increase for each compound with the acetone concentration in the mobile phase. However, above a certain acetone concentration they tend to migrate with the solvent front. The most hydrophilic compounds—carbenecillin and metampicillin—were the first to reach a maximum R_f. On the other hand, at 0% acetone in the mobile phase, the most lipophilic compounds remained near the origin; dicloxacillin did not move at all.

The influence of acetone concentration on the migration of compounds could also be observed in reversed-phase TLC of cephalosporins (Figures 1b, c) (*11*). At 10% acetone in the mobile phase all the compounds showed a greater migration than at 0% acetone. The R_f values are plotted in Figure 3 against the acetone concentration in the mobile phase. Compound I, which appeared to be the most lipophilic, did not move until a certain acetone concentration was reached. On the other hand, Compound XIII migrated with the solvent front at acetone concentrations higher than 2%.

With the following equation:

$$R_m = \log\left(\frac{1}{R_f} - 1\right)$$

the R_f values were transformed into R_m values. Negative and positive R_m values were derived from R_f values greater and smaller than 0.5. The plots of Figures 2 and 3 were transformed into the plots of Figures 4 and 5.

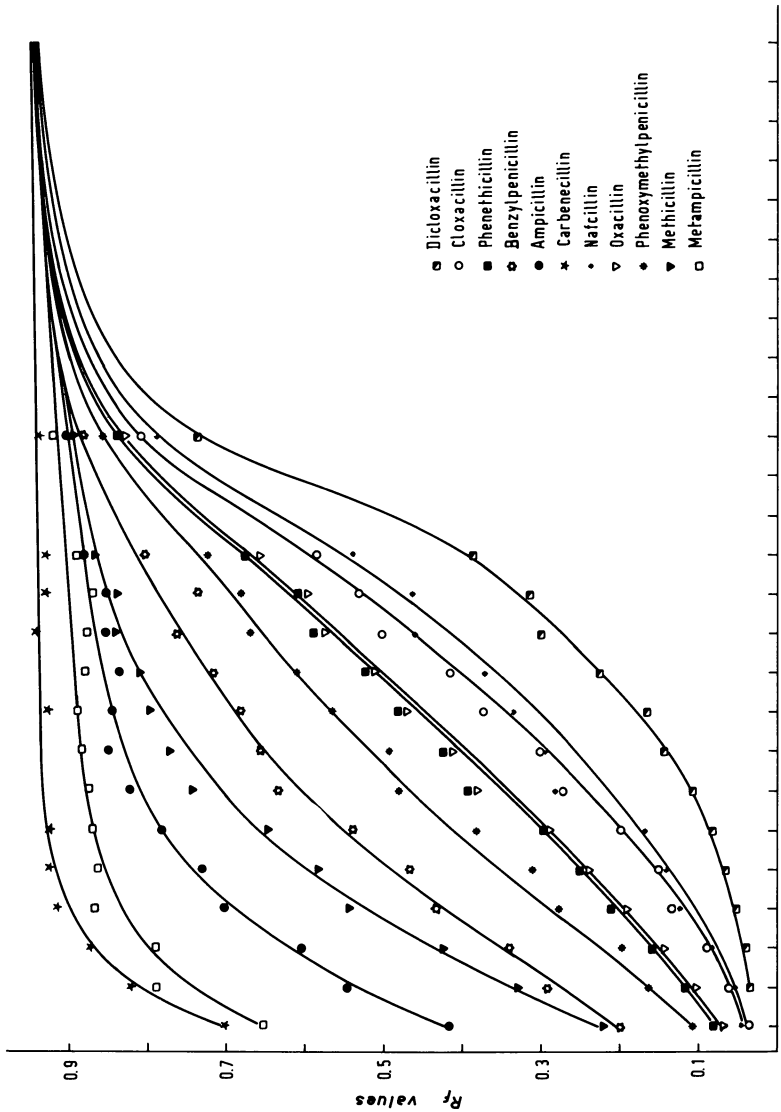


Figure 2. R_f *values of penicillins vs. acetone concentration in the mobile phase*

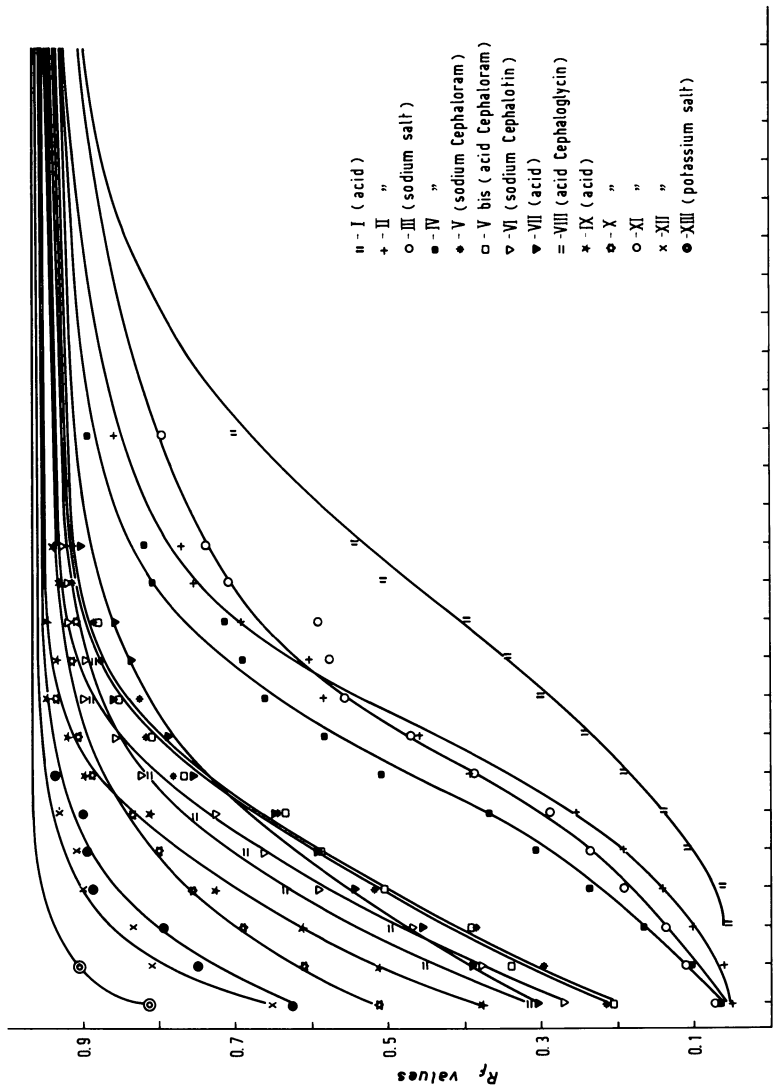


Figure 3. R_f *values of cephalosporins vs. acetone concentration in the mobile phase*

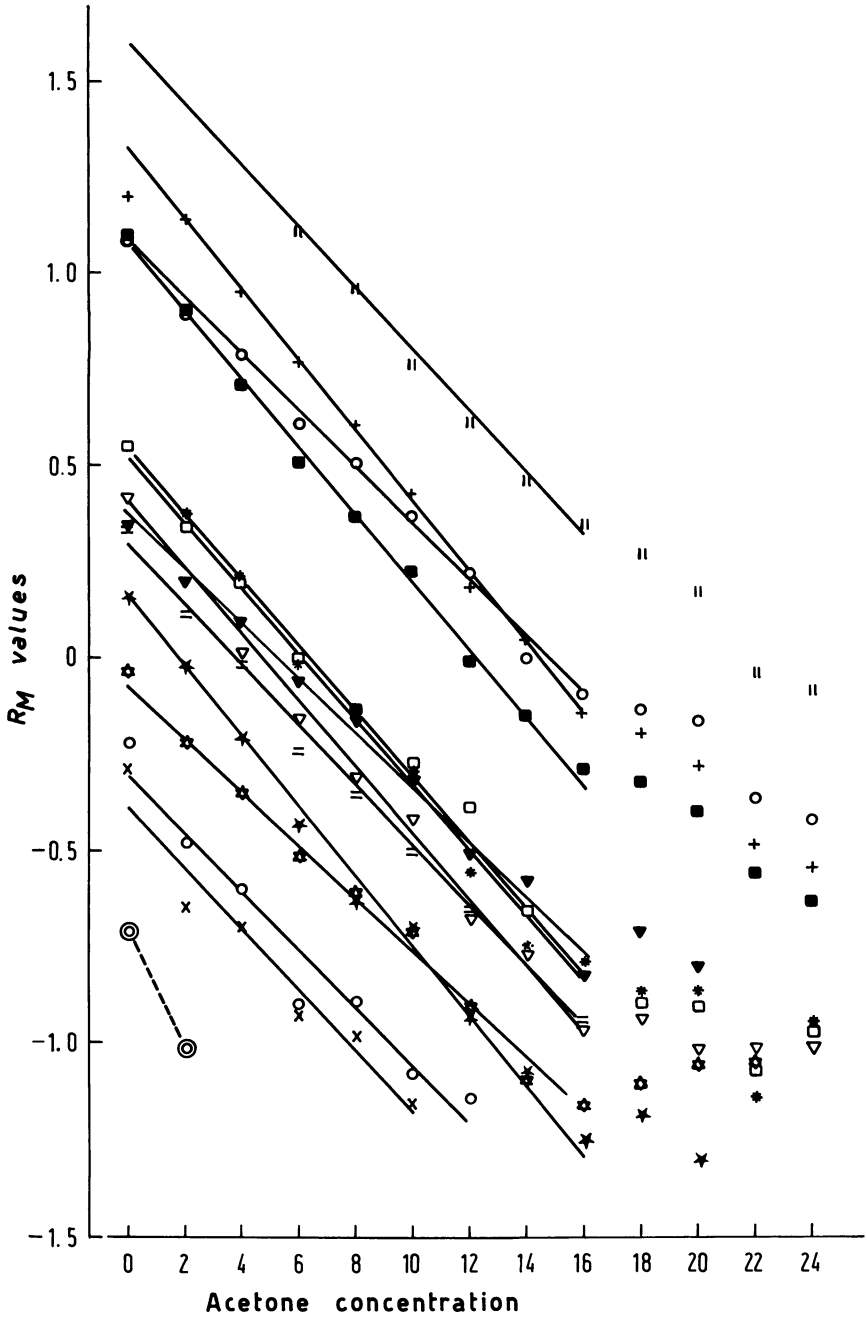


Figure 4. Linear relationship between R_m *values of penicillins and acetone concentration in the mobile phase. The compounds indicated are listed in Figure 2.*

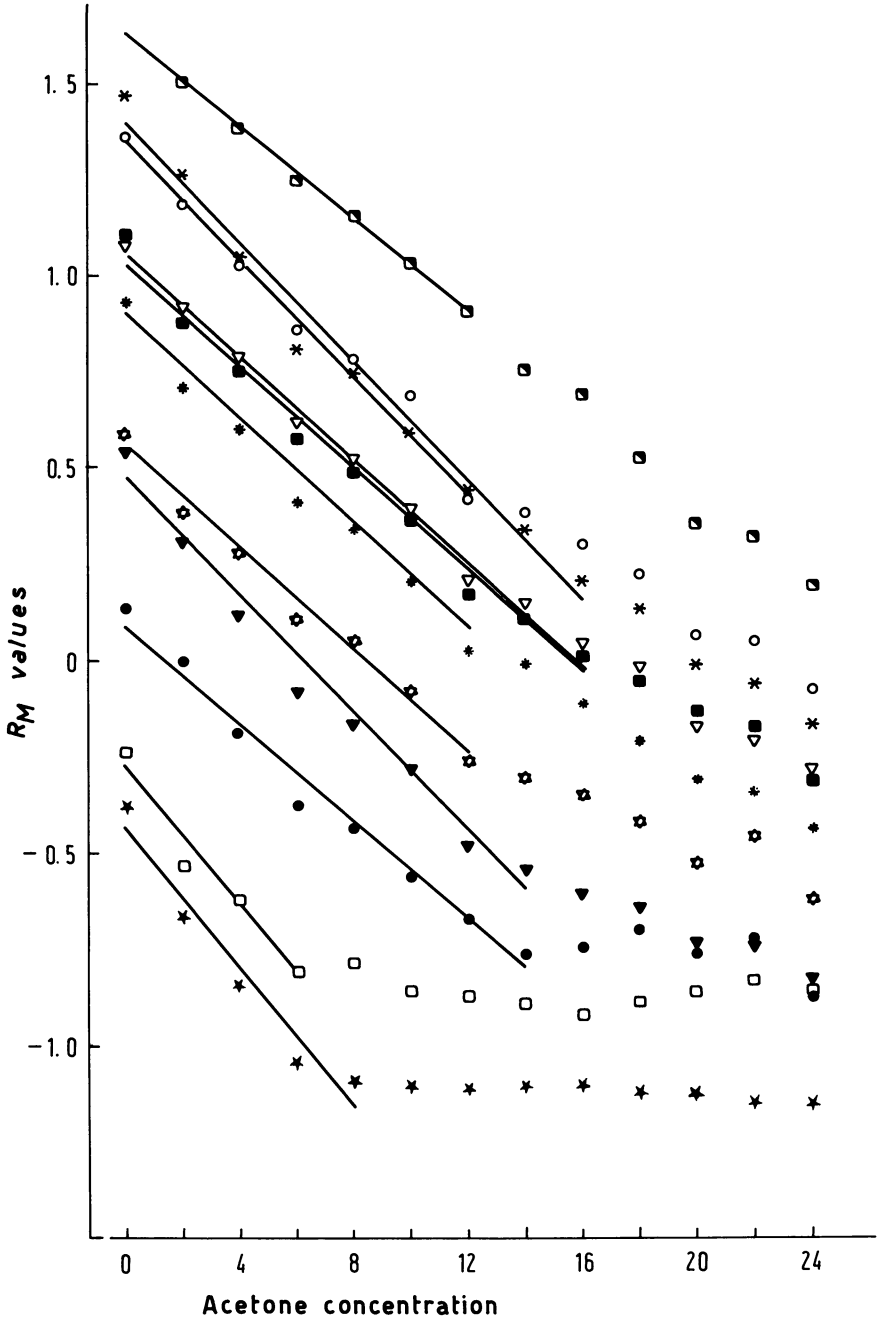
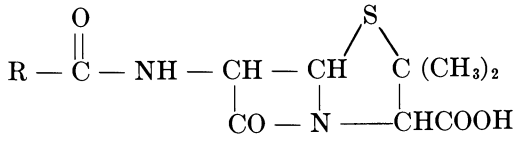
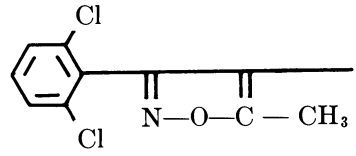
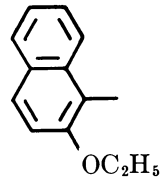
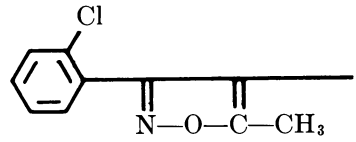
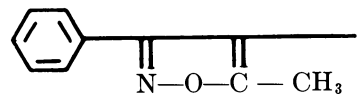
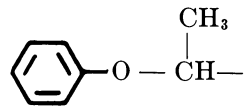
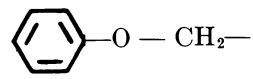
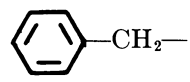


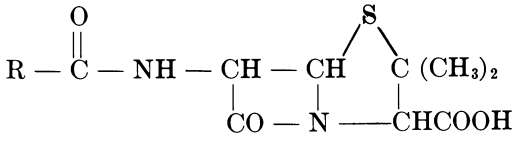
Figure 5. Linear relationship between R_m *values of cephalosporins and acetone concentration in the mobile phase. The compounds indicated are listed in Figure 3.*

Table I. Lipophilic Character of Penicillins



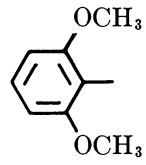
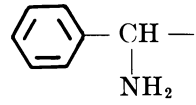
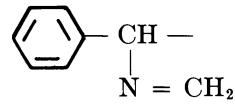
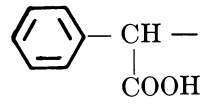
Compound	*R*
Dicloxacillin	
Nafcillin	
Cloxacillin	
Oxacillin	
Phenethicillin	
Phenoxymethylpenicillin	
Benzylpenicillin	

as Expressed by Their R_m and $\Sigma\pi$ Values



R_m	$\Sigma\pi$	*Log* K	*log* (B/F × 10_3)
1.62			
1.39	3.54		3.921
1.34			
1.05			
1.03	2.61	2.28	3.643
0.89	2.11	2.09	3.588
0.55	2.69	1.83	3.188

Table I.

Compound	*R*
Methicillin	
Ampicillin	
Metampicillin	
Carbenicillin	

For each compound there is a range of linear relationships between R_m values and acetone concentration in the mobile phase. The straight lines of Figures 4 and 5 were calculated by the least-squares method from the R_m values in the range of linearity. The range of linear relationship is limited by the fact that above a certain acetone concentration in the mobile phase all the compounds tend to migrate with the solvent front. This is particularly evident in the least lipophilic compounds which tend to reach a maximum R_f with a 2–4% acetone concentration. According to Boyce and Milborrow (*3*) the R_m values in the linearity range were considered to be satisfactory; over this range there are maximum increments of R_m values for each compound and among different compounds. To obtain an R_m value for each compound in a standard system where they all could be compared, the theoretical R_m values at 0% acetone in the mobile phase were calculated by interpolation or extrapolation from the linear part of the curve. The calculations were done using the equations of the straight lines of Figures 4 and 5. The calculated R_m values are reported in Tables I and III for penicillins and

Continued

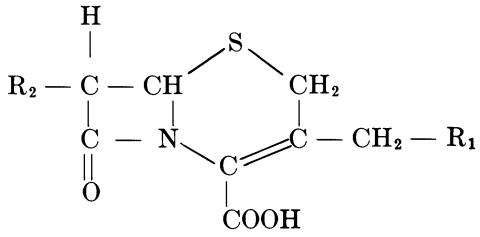
R_m	$\Sigma\pi$	*Log* K	*log* (B/F × 10_3)
0.47	1.47	1.22	2.982
0.07	0.84		2.340
−0.28			
−0.46			

cephalosporins. Table I shows the influence of the substituent group on the partition data of the penicillins. When Cl atoms are introduced into the aromatic ring of oxacillin, lipid solubility increases, as shown by the increased R_m values of cloxacillin and dicloxacillin. When a CH_3 group is introduced into the side chain of phenoxymethylpenicillin, the R_m value of phenethicillin increases. On the other hand, water solubility increases after NH_2, N=CH_2, and COOH groups are introduced into the side chain of benzylpenicillin. Methicillin becomes more hydrophilic than benzylpenicillin as a result of the absence of the benzylmethylene group and the introduction of OCH_3 groups in the 2- and 6-positions. The effect of substituent groups on the partition coefficients of the penicillins agree with the results of Hansch *et al.* (*12*, *13*) and Bird and Marshall (*14*).

Table II. Regression Analysis of the Data of Table I

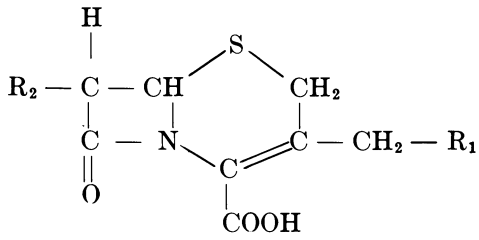
	Equation	n	r	s
1.	$\log B/F = 2.196 + 2.042\, R_m - 0.580\, R_m^2$	6	0.998	0.048
2.	$\log B/F = 1.543 + 1.154\, \Sigma\pi - 0.144\, \Sigma\pi^2$	6	0.916	0.294
3.	$R_m = -0.225 + 0.434\, \Sigma\pi$	6	0.892	0.236

Table III. Lipophilic Character of Cephalosporins



Compound	R_1
I (acid)	O—CO—CH_3
II (acid)	O—CO—CH_3
III (sodium salt)	N_3
IV (sodium salt)	O—CO—CH_3
V (sodium cephaloram)	O—CO—CH_3
VI (sodium cephalotin)	O—CO—CH_3
VII (acid)	N_3
VIII (acid cephaloglycin)	O—CO—CH_3
IX (acid)	O—CO—CH_3
X (acid)	OH

as Expressed by Their R_m Values



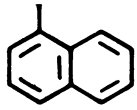
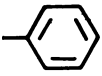
R_2	R_m *value*
NH—CO • $(CH_2)_6$ • CH_3	1.60
NH — CO — CH_2 	1.32
NH — CO — CH_2S — CH_2 	1.08
NH — CO — CH_2S — CH_2 	1.08
NH — CO — CH_2 	0.54
NH — CO — CH_2 	0.40
NH — CO — CH_2 	0.37
NH — CO — CH(NH$_2$) 	0.29
NH — CO — CH_2 	0.16
NH — CO — CH_2 	−0.07

Table III.

Compound	R_1
XI (acid)	O—CO—CH_3
XII (7-aminocephalosporanic acid)	O—CO—CH_3
XIII (potassium cephalosporin C)	O—CO—CH_3

They used the partition coefficient function π and were able to show the hydrophobic character of substituents such as the CH_3 group in the side chain of phenoxymethylpenicillin or Cl atoms in the para positions of the aromatic ring. The $\Sigma\pi$ and log K values calculated by Bird and Marshall (*14*) for some penicillins are listed in Table I along with the log (B/F) values, indicating the serum binding of penicillins (*14*). A regression analysis of the data of Table I is reported in Table II. The serum binding of penicillins shows a parabolic dependence on lipophilic character. However, the correlation coefficient obtained with the R_m values (Equation 1) is higher than that provided by the $\Sigma\pi$ values (Equation 2). This is reflected in Equation 3 which shows that the correlation between R_m and $\Sigma\pi$ values is not very high. This seems to be the result of the disagreement between the R_m and the $\Sigma\pi$ values of benzylpenicillin. The calculated $\Sigma\pi$ value for benzylpenicillin indicates a lipophilic character higher than that expressed by its R_m and log K values. In fact, there seems to be a closer correlation between the R_m value and the logarithm of the experimental octanol–water partition coefficients. This would suggest that TLC might avoid the anomalies observed in the $\Sigma\pi$ values when they are compared with the experimental log P values.

The calculated R_m values in Table III show the influence of substituent groups on the lipophilic character of cephalosporins. The R_m value of Compound II decreases more and more with the substitution of the naphthyl group by a benzene, a thiophene, or a furan ring as in Compounds V, VI, and IX. The hydrophilic character of V increases when an NH_2 group is introduced into the side chain (Compound VIII) or when the $OCOCH_3$ group is replaced by an OH (Compound X), or when the benzene ring is replaced by a Cl atom (Compound XI).

Steroids. The same kind of results were obtained with steroids (*15*). However, to obtain suitable migrations, it was necessary to use higher concentrations of acetone or methanol in the mobile phase. Plots of R_m values of testosterone esters *vs.* acetone or methanol concentration in the

Continued

R_2	R_m *value*
$NH—CO—CH_2Cl$	−0.31
NH_3^+	−0.39
$NH—CO\ (CH_2)_3—CH(NH_2)(COOH)$	−0.71

mobile phase are reported in Figures 6 and 7. The equations of the straight lines of Figures 6 and 7 were used to calculate an R_m value corresponding to a 54% concentration of acetone or methanol in the mobile phase. The calculated R_m values are listed in Table IV, where the π values (provided by Hansch) are also reported. The π value of the phenyl propionate ester indicates a lipophilic character higher than that expressed by its R_m value. The $C_6H_5CH_2$ group of the phenyl propionate ester is also present in the side chain of benzylpenicillin, where a similar disagreement between π and R_m values was observed. This could be explained on the basis of some group interactions. However, insufficient data are available for any conclusion on this point. A regression analysis of the data of Table IV is reported in Table V, where Equation 1 shows a significant relationship between R_m values calculated at 54% acetone and those calculated at 54% methanol. This result agrees with those of Collander; he showed a linear relationship between the partition coefficients obtained with two different sets of solvents (*16*). Equations 2 and 3 show a good correlation between both sets of R_m values and π values (Table IV). Iwasa *et al.* (*17*) also found a good correlation between π and R_m values of phenols in further support of Collander's findings.

The suitability of the silicone oil–aqueous acetone system for measuring R_m values for correlation with biological activity should also be supported by Collander's results (*10, 16, 18*). On the basis of his work, Collander pointed out that the nature of the phase should not affect the results qualitatively. He found that ether–water and olive oil–water partition coefficients are correlated equally with penetration into *Nitella* cells (*19*). Boyce and Milborrow (*3*) pointed out a parabolic relationship between the molluscicidal activity of *N*-*n*-alkyltritylamines and the R_m values obtained with a chromatographic system of liquid paraffin and 70% acetone–water. With the present chromatographic system good correlations between R_m values and biological activity were found for bisdichloroacetamides (*9, 20*), vitamin K analogs (*9, 20*), penicillins (*21*), cephalosporins (*21*), and testosterone esters (*22*).

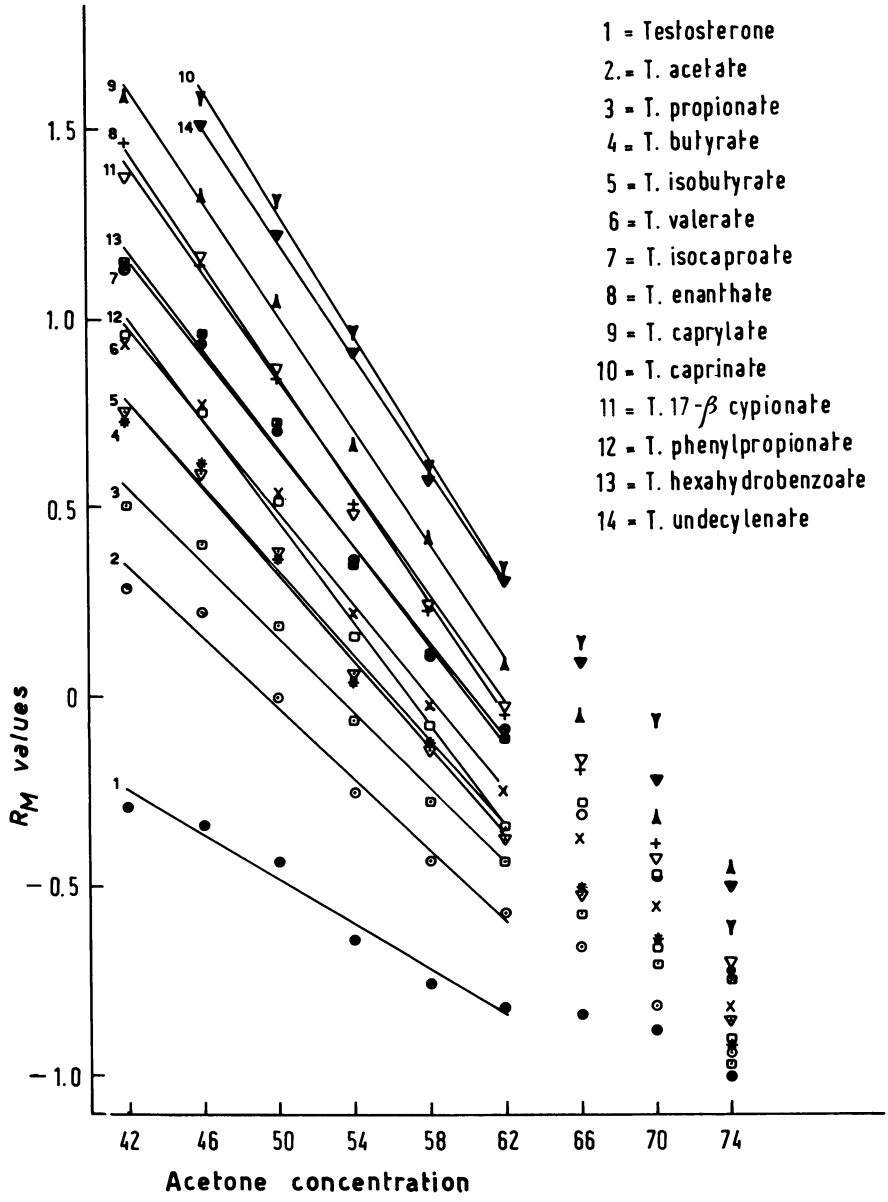


Figure 6. Linear relationship between R_m values of testosterone derivatives and acetone concentration in the mobile phase

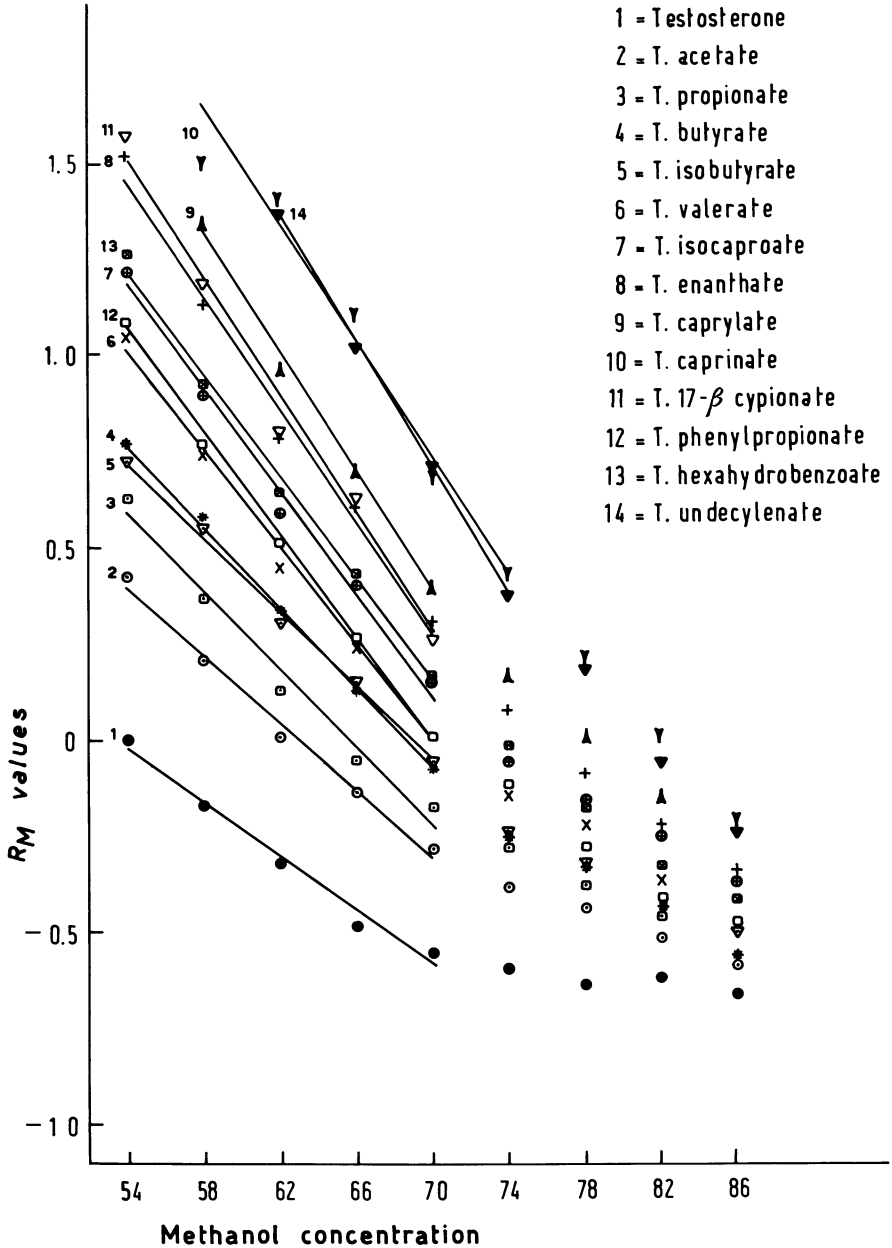
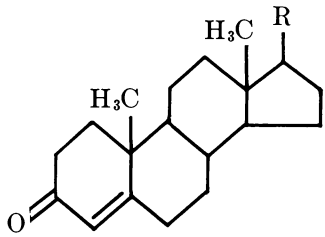


Figure 7. Linear relationship between R_m values of testosterone derivatives and methanol concentration in the mobile phase

Table IV. Lipophilic Character of Testosterone and Its Derivatives as Expressed by the π and R_m Values



Compound	*R*	π *values*	R_m *values* (Me_2CO)	R_m *values* (M_eOH)
Testosterone	OH	−1.160	−0.600	−0.020
-acetate	$OCOCH_3$	−0.270	−0.220	0.400
-propionate	$OCOC_2H_5$	0.230	−0.050	0.590
-butyrate	$OCO(CH_2)_2CH_3$	0.730	0.090	0.770
-isobutyrate	$OCOCH(CH_3)_2$	0.530	0.090	0.730
-valerate	$OCO(CH_2)_3CH_3$	1.230	0.250	1.020
-isocaproate	$OCO(CH_2)_2CH(CH_3)_2$	1.530	0.400	1.190
-enanthate	$OCO(CH_2)_5CH_3$	2.230	0.560	1.450
-caprylate	$OCO(CH_2)_6CH_3$	2.730	0.700	1.630
-caprinate	$OCO(CH_2)_8CH_3$	3.730	0.960	1.960
-17-β-cypionate	$OCOCH_2CH_2$ 	2.370	0.540	1.520
-phenylpropionate	$OCOCH_2CH_2C_6H_5$	2.360	0.190	1.060
-hexahydrobenzoate	$OCOC_6H_{11}$	1.740	0.400	1.230
-undecylenate	$OCO(CH_2)_8CH{=}CH_2$	3.930	0.900	2.040

Table V. Regression Analysis of the Data of Table IV

Equation	n	r	s
1. $R_m(Me_2CO) = -0.509 + 0.728\,R_m(MeOH)$	14	0.993	0.051
2. $R_m(Me_2CO) = -0.143 + 0.288\,\pi$	14	0.964	0.119
3. $R_m(MeOH) = -0.496 + 0.394\,\pi$	14	0.981	0.118

What are the disadvantages and advantages of the chromatographic method *vs.* the experimental method for determining the partition coefficient? The experimental method for determining the octanol–water partition coefficient is likely to give more accurate and unequivocal data. However, the chromatographic method has several advantages (*3*, *4*): (a) it is simple and rapid; (b) it requires little material; (c) the material does not need to be very pure because impurities are separated during the determination; (d) the detection of spots by unspecific methods

avoids the need for specific quantitative analytical methods; (e) the determination of the partition coefficient of slightly water-soluble compounds requires a long period of equilibration to achieve thorough partitioning between the phases; (f) the existence of a range of linearity between R_m values and mobile phase composition allows one to obtain an R_m in a standard system for each compound in a series.

TLC seems to be a suitable technique for determining partition data of antibiotics and steroids. Future determinations should provide further evidence of the relationship between π and R_m values.

Literature Cited

1. Daniels, T. C., Jorgensen, E. C., "Textbook of Organic, Medicinal, and Pharmaceutical Chemistry," C. O. Wilson, O. Gisvold, R. F. Doerge, Eds., p. 4 ff., Lippincott, Philadelphia, 1966.
2. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **86,** 1616.
3. Hansch, C., Steward, A. R., Iwasa, J., Deutsch, E. W., *Mol. Pharmacol.* (1965) **1,** 205.
4. Boyce, C. B. C., Milborrow, B. V., *Nature* (1965) **208,** 537.
5. Bate-Smith, E. C., Westall, R. G., *Biochim. Biophys. Acta* (1950) **4,** 427.
6. Consden, R., Gordon, A. H., Martin, A. J. P., *Biochem. J.* (1944) **38,** 224.
7. Martin, A. J. P., *Biochem. Soc. Symp. (Cambridge, England)* (1950) **3,** 4.
8. Green, J., Marcinkewicz, S., *J. Chromatog.* (1963) **10,** 35.
9. Biagi, G. L., Merola, A. J., Cornwell, D. G., *Federation Proc. 51st Ann. Meetg., Chicago* (1967).
10. Biagi, G. L., Barbaro, A. M., Gamba, M. F., Guerra, M. C., *J. Chromatog.* (1969) **41,** 371.
11. Biagi, G. L., Barbaro, A. M., Guerra, M. C., Gamba, M. F., *J. Chromatog.* (1969) **44,** 195.
12. Hansch, C., Steward, A. R., *J. Med. Chem.* (1964) **7,** 691.
13. Hansch, C., Deutsch, E. V., *J. Med. Chem.* (1965) **8,** 705.
14. Bird, A. E., Marshall, A. C., *Biochem. Pharmacol.* (1967) **16,** 2275.
15. Biagi, G. L., Guerra, M. C., Barbaro, A. M., *J. Med. Chem.* (1970) **13,** 944.
16. Collander, R., *Acta Physiol. Scand.* (1954) **7,** 420.
17. Iwasa, J., Fujita, T., Hansch, C., *J. Med. Chem.* (1965) **8,** 150.
18. Collander, R., *Acta Chem. Scand.* (1951) **5,** 774.
19. Collander, R., *Physiol. Plantarum* (1954) **7,** 420.
20. Turnbull, J. D., Biagi, G. L., Merola, A. J., Cornwell, D. G., *Biochem. Pharmacol.* (1971) **20,** 1383.
21. Biagi, G. L., Guerra, M. C., Barbaro, A. M., Gamba, M. F., *J. Med. Chem.* (1970) **13,** 511.
22. Biagi, G. L., Guerra, M. C., Barbaro, A. M., *J. Med. Chem.* (1970) **13,** 944.

RECEIVED May 10, 1971.

6

Substituent-Effect Analyses of the Rates of Metabolism and Excretion of Sulfonamide Drugs

TOSHIO FUJITA

Department of Agricultural Chemistry, Kyoto University, Kyoto, Japan

Analyses of the rates of metabolism and renal excretion of sulfonamide drugs were attempted in terms of their physicochemical properties. Using simple models for acetylation in the liver and excretion from the kidney, expressions for rate constants of both processes were proposed on the basis of free energy related substituent constants. Taking into account the effect of dissociation, logarithmic values of the rate constants correlated with a linear combination of the hydrophobicity constant, π, and the electronegativity parameter, ΔpK_A, of the N-1 substituent of sulfanilamide. The information obtained from the correlations should be useful in designing long-acting sulfonamide drugs.

Previously, we correlated bacteriostatic activity and structure of various series of N-1 substituted sulfonamide drugs with the free energy related substituent constants of the N-1 substituent (*1*). For a series of N-1 heterocyclic derivatives, activity data obtained by Krüger-Thiemer and Bünger (*3*) were well correlated by Equation 1. In Equation 1, C

$$\log \frac{1}{C} + \log \frac{K_A + [H^+]}{[H^+]} = -0.296\pi^2 + 0.985\pi + 0.605\Delta pK_A - 2.090 \quad (1)$$
$$n = 17 \qquad r = 0.975 \qquad s = 0.223$$

is the minimum inhibitory concentration in μmole/liter, π is the hydrophobicity constant of the N-1 substituent derived from the partition coefficient, P, with an isobutyl alcohol–water system as $\pi = \log P$ (sub-

stituted sulfanilamide) − log P_o (sulfanilamide), and ΔpK_A is the electronegativity parameter of the substituent defined as $\Delta pK_A = pK_A^\circ$ (sulfanilamide) − pK_A (substituted sulfanilamide). The n is the number of points used in the regression, r is the correlation coefficient, and s is the standard deviation. Although the second term on the left is introduced to correct bacteriostatic activity according to the extent of dissociation at the experimental pH, Equation 1 does not necessarily mean that the molecular species responsible for the activity is the neutral form. The correlation was quite good, and information on the optimum pK_A and π values required for the maximum activity was obtained by taking partial differentials of this equation.

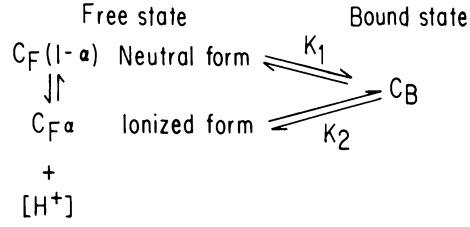


Figure 1. Model for protein binding

We also analyzed the binding constant of sulfonamides to human plasma protein (*1*). In the free, non-bound state the drugs exist as an equilibrium mixture of neutral and ionized forms, of which the concentrations are $C_F(1 - \alpha)$ and $C_F\alpha$, respectively (*see* Figure 1). In the bound state they exist as only one molecular species, C_B. The effective binding constant, K, can be expressed by the respective binding constants of neutral and ionized forms: K_1 and K_2, as shown in Equation 2. Taking

$$K = \frac{C_B}{C_F} = \frac{K_1 K_2}{K_1 + K_2} \tag{2}$$

into account the dissociation equilibrium of free drugs and assuming that the logarithmic values of the fundamental binding constants, K_1 and K_2, can be described by free energy related parameters, we derived Equation 3 for protein binding; where a, ρ, and c are constants which are determined

$$\log K + \log \frac{K_A + [H^+]}{[H^+]} = a\pi + \rho\Delta pK_A + c \tag{3}$$

by the method of least squares. For 20 N-1 heterocyclic sulfonamides, the binding data obtained by Rieder (*4*) were best represented by Equation 4, which shows that only the hydrophobicity of the N-1 sub-

$$\log K + \log \frac{K_A + [H^+]}{[H^+]} = 1.651\pi - 2.896 \quad (4)$$

$$n = 20 \qquad r = 0.938 \qquad s = 0.365$$

stituent plays a significant role in governing the modified binding constant. Equation 3 was also useful in elucidating the serum binding of N-4-acetylsulfanilamides, the major metabolites of sulfonamide drugs, as shown in Equation 5 (*2*).

$$\log K + \log \frac{K_A + [H^+]}{[H^+]} = \underset{(\pm 0.651)}{2.029\pi} - \underset{(\pm 1.126)}{2.860} \quad (5)$$

$$n = 5 \qquad r = 0.985 \qquad s = 0.264$$

In Equations 4 and 5, K is the mass-action equilibrium constant of binding expressed in liters/μmole (*4*). The figures in parentheses are the 95% confidence intervals.

The serum protein binding of sulfonamide drugs has been considered an important factor in maintaining a high level of the drug in the blood since elimination of drugs from blood occurs *via* the non-bound free molecule. The time of duration in blood or the rate of drug elimination (antibacterial activity notwithstanding) is one of the most important properties to be considered in designing new drugs. Encouraged by the correlations obtained for bacteriostatic activity and protein binding shown above, we attempted to analyze the rate of elimination for sulfonamide drugs in terms of free energy related substituent constants.

Methods

According to Nelson (*5*) the rate of elimination of drugs in blood follows the first-order kinetics; the rate constant of which is k_{El}, as shown in Equation 6, where C is the total drug concentration in blood. Sulfona-

$$-\frac{dC}{dt} = k_{El}C = (k_{Me} + k_{Ex})C \doteqdot (k_{Ac} + k_{Ex})C \quad (6)$$

mide drugs are eliminated mainly by two routes, one of which is metabolism in the liver, and the other is renal excretion. The rate of hepatic metabolism and that of renal excretion also obey first-order kinetics so that the rate constant of elimination, k_{El}, comprises the respective first-order rate constants, k_{Me} and k_{Ex}. These constants are estimated by measuring the time course of drug concentration in the blood after administration and the ratio of integrated amounts of metabolized and free drug in the urine. For most sulfonamide drugs, the major metabolite

is the N-4-acetyl derivative; thus, we can set the rate constant of hepatic metabolism, k_{Me}, equal to that of acetylation, k_{Ac}, as shown in Equation 6.

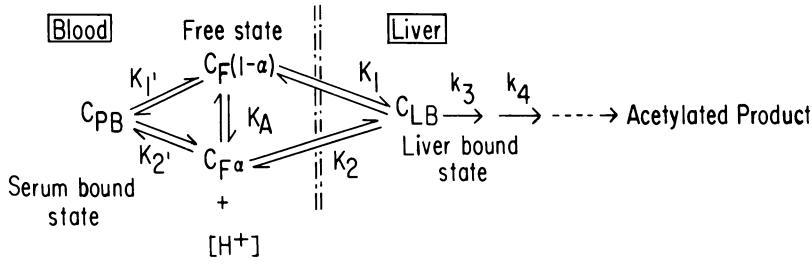


Figure 2. Model of hepatic acetylation process

To correlate the rate constant of hepatic acetylation with free energy related parameters, we used the simple model shown in Figure 2. In blood, the free undissociated and the free dissociated species are in a dynamic quilibrium with each other and also with the serum protein-bound state. Both the neutral and ionized forms of free drug are also in an equilibrium with a liver tissue bound state. We assume that these binding equilibria are established rather quickly, that there is a critical rate-determining step, k_3 after the drug is bound to liver tissue, and that a steady state in the concentration of the liver bound complex, C_{LB}, is maintained. If we apply Equation 2 to this model, the concentrations of serum-bound, liver-bound, and free drug, C_{PB}, C_{LB}, and C_F, are related to each other by Equations 7 and 8, where K_1, K_2, K_1', and K_2' are the binding constants for the unit processes as shown in Figure 2. From

$$C_{PB} = \frac{K_1'K_2'}{K_1' + K_2'} C_F \tag{7}$$

$$C_{LB} = \frac{K_1K_2}{K_1 + K_2} C_F \tag{8}$$

$$-\frac{dC_{PB}}{dt} = -\frac{K_1'K_2'}{K_1' + K_2'} \frac{dC_F}{dt} \tag{9}$$

$$-\frac{dC_F}{dt} = \frac{K_1K_2}{K_1 + K_2} k_3C_F \tag{10}$$

Equation 7, we can derive Equation 9 for the rate of decrease of C_{PB}. Since the rate of decrease of C_F can be expressed by Equation 10, the rate of decrease of the total concentration in the blood, $(C_F + C_{PB})$, is described by Equation 11, and the rate constant of acetylation, k_{Ac}, is derived as shown in Equation 12.

$$-\frac{dC}{dt} = -\frac{d(C_{PB} + C_F)}{dt} = -\frac{dC_F}{dt}\left(1 + \frac{K_1'K_2'}{K_1' + K_2'}\right)$$
$$= C_F\left(1 + \frac{K_1'K_2'}{K_1' + K_2'}\right)\frac{K_1K_2}{K_1 + K_2}k_3 = C\frac{K_1K_2}{K_1 + K_2}k_3 \quad (11)$$

$$k_{Ac} = \frac{K_1K_2}{K_1 + K_2}k_3 \quad (12)$$

For the logarithmic value of k_{Ac}, Equation 13 and its modification, Equation 14, were formulated, where $K_A°$ is the dissociation constant of sulfanilamide, a, ρ', c', ρ, and c are constants, and $[H^+]$ is the hydrogen ion concentration in the blood. We used Equation 14 to analyze the acetylation data. The second term on the left of Equation 14 expresses the dissociation effect of the free drug, but it does not mean that the neutral form is only responsible for metabolism.

$$\log k_{Ac} + \log\frac{[H^+] + K_A}{[H^+] + K°_A} = a\pi + \rho'\,\Delta pK_A + c' \quad (13)$$

$$\log k_{Ac} + \log\frac{[H^+] + K_A}{[H^+]} = a\pi + \rho\,\Delta pK_A + c \quad (14)$$

These equations are derived in a manner similar to that for Equation 3, as previously reported (*1*), assuming that log k_3 can be described by a linear combination of the free energy related substituent constants.

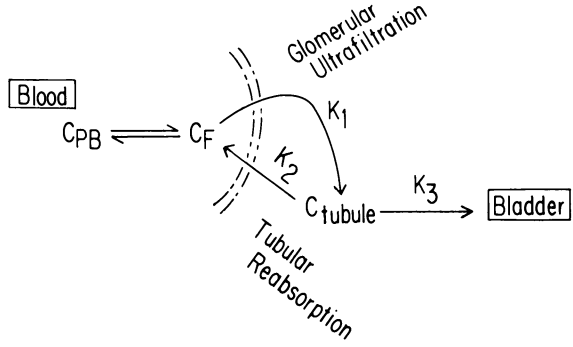


Figure 3. Model of renal excretion

For renal excretion, we used the model shown in Figure 3. Here, the rate constant for the decrease in the concentration of drug in the blood by glomerular ultrafiltration is k_1, for the tubular reabsorption process is k_2, and for the movement of drug from tubule to bladder is k_3. We assume that k_2 is much larger than k_3 and that concentration of the

drug in the tubule (C_{tubule}) is in a steady state. Thus, the rate constant for the whole process of renal excretion can be expressed by Equation 15.

$$k_{Ex} = \frac{k_1 k_3}{k_2} \tag{15}$$

We further assume that the total volume of blood, V, is much larger than that filtered per unit time, dv. The rate of decrease for the amount of free drug *via* glomerular ultrafiltration is $C_F dv$ at any time so that the rate of decrease for the concentration in blood ($-(dC_F/dt)_G$) is given as Equation 16. Therefore, the rate of decrease for the C_{total} by the same route can be described by Equation 17. Thus, the rate constant, k_1, is

$$-\left(\frac{dC_F}{dt}\right)_G = C_F \frac{dv}{V} \tag{16}$$

$$-\left(\frac{dC}{dt}\right)_G = -\left(\frac{dC_{PB}}{dt} + \frac{dC_F}{dt}\right)_G = -\left(1 + \frac{K_1' K_2'}{K_1' + K_2'}\right)\left(\frac{dC_F}{dt}\right)_G$$

$$= \left(1 + \frac{K_1' K_2'}{K_1' + K_2'}\right) C_F \frac{dv}{V} = C \frac{dv}{V} \tag{17}$$

described as dv/V which is constant through a series of compounds; unless the test organism is changed. Since the ratio of k_2 to k_3 would be proportional to the concentration of the neutral drug in the tubule and also to its permeability through the tubular membrane, the logarithm of k_2/k_3 can be described as Equation 18, where $[H^+]$ is that of urine in

$$\log \frac{k_2}{k_3} = \log \frac{[H^+]}{K_A + [H^+]} + \log \text{perm.} + \text{constant} \tag{18}$$

the tubule. Log perm. can be further related to the free energy related parameters; thus, the value of log k_{Ex} is expressed by Equation 19 which can be modified to Equation 20, where a, ρ, and c are constants.

$$\log k_{Ex} = \log k_1 - \log \frac{k_2}{k_3} = \log \frac{K_A + [H^+]}{[H^+]} + a\pi + \rho \Delta \text{p}K_A + c \tag{19}$$

$$\log k_{Ex} - \log \frac{K_A + [H^+]}{[H^+]} = a\pi + \rho \Delta \text{p}K_A + c \tag{20}$$

Using Equations 14 and 20, we analyzed the data shown in Tables I and II obtained by Nogami (*6*), Yamazaki (*7*), and Kakemi (*8, 9*) and their associates by the method of least squares. We have not included all compounds of the original set in the analyses since the physicochemical parameters of some derivatives are lacking.

Table I. Substituent Constants

			Rat		
				Acetylation Rate[h]	
Compounds	ΔpK_A[b]	π[c]	*log* k_{Ac}	*obs.*	*calc.*[d]
Sulfanilamide	0.00	0.00	−0.45	−0.45	−1.10
Acetosulfamine	5.05	1.01	−0.36	1.64	—
Sulfathiazole	3.35	0.82	−1.06	−0.58	−0.43
Sulfadiazine	4.30	1.11	−1.47	−0.20	−0.19
Sulfamerazine	3.52	0.94	−1.33	−0.73	−0.33
Sulfisoxazole	5.83	2.61	−1.21	1.57	1.04
Sulfisomidine	3.07	0.39	−0.96	−0.65	−0.78
Sulfaphenazole	4.54	2.15	−1.00	0.50	0.66
Sulfamethoxypyridazine	3.40	1.02	−1.42	−0.91	−0.26
Sulfadimethoxine	4.40	1.86	−1.92	−0.55	—
Sulfisomezole	4.64	1.55	−1.38	0.22	0.17
Sulfamonomethoxine	4.42	1.37	−1.36	0.03	0.02

[a] The rate constant, k_{Ac}, is the value per hour.
[b] Derived from the values of pK_A in Ref. *16*.
[c] Calculated from the "Übergangszahlen" in Ref. *3* with correction for ionization in the aqueous phase according to the K_A values obtained by Yamazaki *et al.*, Ref. *16*.
[d] Calculated by Equation 28.

Table II. Substituent Constants

				Rat		
					Excretion Rate[h]	
Compounds	ΔpK_A[b]	π[b]	π'[c]	*log* k_{Ex}	*obs.*	*calc.*[d]
Sulfanilamide	0.00	0.00	0.00	−1.18	−1.18	−1.06
Acetosulfamine	5.05	1.01	1.61	−0.19	−1.23	−1.17
Sulfathiazole	3.35	0.82	0.68	−0.63	−0.71	−1.28
Sulfadiazine	4.30	1.11	1.81	−1.14	−1.58	−1.41
Sulfamerazine	3.52	0.94	1.84	−1.17	−1.28	−1.37
Sulfisoxazole	5.83	2.61	2.56	−0.61	−2.39	−2.62
Sulfisomidine	3.07	0.39	0.89	−0.64	−0.69	−0.91
Sulfaphenazole	4.54	2.15	2.91	−1.55	−2.16	−2.39
Sulfamethoxypyridazine	3.40	1.02	2.14	−1.55	−1.64	−1.47
Sulfadimethoxine	4.40	1.86	3.12	−1.92	−2.43	−2.13
Sulfisomezole	4.64	1.55	1.97	−1.34	−2.03	−1.78
Sulfamonomethoxine	4.42	1.37	2.13	−1.41	−1.93	−1.64

[a] The rate constant, k_{Ex}, is the value per hour.
[b] The same values as in Table I.
[c] Calculated from the values of partition coefficient measured using a system of $CHCl_3$–pH 7.4 buffer solution (*16*).
[d] Calculated by Equation 38.

and Rate of Acetylation[a]

Rat			*Rabbit*			*Human*		
	Acetylation Rate[h]			*Acetylation Rate*[h]			*Acetylation Rate*[h]	
log k_{Ac}	*obs.*	*calc.*[e]	*log* k_{Ac}	*obs.*	*calc.*[f]	*log* k_{Ac}	*obs.*	*calc.*[g]
−1.41	−1.41	−1.26	−0.92	−0.92	−1.01	−1.45	−1.45	−1.81
—	—	—	−0.90	1.10	—	−1.58	0.42	—
—	—	—	−0.47	0.01	−0.06	−1.29	−0.81	−0.89
−1.51	−0.24	−0.28	−0.50	0.77	0.28	−1.51	−0.24	−0.57
—	—	—	−0.26	0.34	0.08	−1.26	−0.66	−0.76
−1.72	1.06	1.03	−0.87	1.91	2.02	−1.40	1.38	1.12
−0.96	−0.65	−0.91	−1.66	−1.35	−0.56	−2.27	−1.96	−1.38
—	—	—	−0.58	0.92	1.49	−1.46	0.04	0.60
—	—	—	−0.61	−0.10	0.17	−1.57	−1.06	−0.67
−2.28	−0.91	—	−1.16	0.21	—	−2.52	−1.15	—
−1.38	0.22	0.10	−0.24	1.36	0.79	−1.25	0.35	−0.07
−1.74	−0.35	−0.06	−0.55	0.84	0.58	−1.68	−0.29	−0.27

[e] Calculated by Equation 22.
[f] Calculated by Equation 31.
[g] Calculated by Equation 34.
[h] The acetylation rate is obtained from: $\log k_{Ac} + \log (K_A + [H^+])/([H^+])$. (*See* Equation 14).

and Rate of Excretion[a]

Rat			*Rabbit*			*Human*		
	Excretion Rate[h]			*Excretion Rate*[h]			*Excretion Rate*[h]	
log k_{Ex}	*obs.*	*calc.*[e]	*log* k_{Ex}	*obs.*	*calc.*[f]	*log* k_{Ex}	*obs.*	*calc.*[g]
−1.82	−1.82	−1.81	−1.02	−1.03	−0.86	−1.50	−1.50	−1.51
—	—	—	−0.28	−3.68	−3.49	−1.45	−2.15	—
—	—	—	−0.36	−2.07	−2.68	−1.05	−1.08	−1.30
−1.16	−1.60	−1.62	−0.71	−3.36	−3.23	−1.30	−1.54	−1.70
—	—	—	−0.68	−2.56	−2.81	−1.90	−1.95	−1.84
−0.70	−2.48	−2.65	−0.26	−4.44	−4.63	−1.11	−2.52	—
−1.02	−1.07	−1.28	−0.92	−2.36	−2.34	−1.08	−1.10	—
—	—	—	−0.72	−3.61	−3.86	−1.64	−1.99	−2.19
—	—	—	−0.94	−2.70	−2.80	−2.17	−2.21	−2.00
−1.84	−2.35	−2.34	−1.52	−4.27	−3.65	−1.84	−2.12	−2.32
−1.34	−2.03	−1.96	−0.65	−3.64	−3.59	−1.39	−1.80	−1.72
−1.63	−2.15	−1.85	−0.87	−3.64	−3.41	−1.93	−2.22	−1.83

[e] Calculated by Equation 44.
[f] Calculated by Equation 47.
[g] Calculated by Equation 52.
[h] The excretion rate is obtained from: $\log k_{Ex} - \log (K_A + [H^+])/([H^+])$. (*See* Equation 20).

Results and Discussion

For the rate of acetylation of six sulfonamide drugs in the rat (*6*) the highest correlation was obtained in Equation 22. If we neglect the second term on the left of the equation and analyze the log k_{Ac} values directly with the substituent parameters, the correlation becomes poorer as shown in Equation 25. For data obtained independently on 10 drugs (*7*), Equation 28 shows the best correlation. An F test shows that the ΔpK_A term in Equation 29 is justified only at the 0.75 level of significance. For 10 drugs used with rabbit (*7*) and human (*8*), Equations 31 and 34 show the best correlation.

Rat (6) ($[H^+] = 10^{-7.4}$)

	n	s	r	
$\log k_{Ac} + \log \frac{K_A + [H^+]}{[H^+]}$				
$= -1.631 + 0.378\Delta pK_A$	6	0.366	0.919	(21)
$= -1.255 + 0.876\pi$ $(\pm 0.432)\ (\pm 0.299)$	6	0.222	0.971	(22)
$= -1.395 + 0.679\pi + 0.100\Delta pK_A$	6	0.229	0.977	(23)
$\log k_{Ac}$				
$= -1.233 - 0.059\Delta pK_A$	6	0.290	0.420	(24)
$= -1.220 - 0.200\pi$	6	0.244	0.644	(25)
$= -1.382 + 0.116\Delta pK_A - 0.428\pi$	6	0.248	0.738	(26)

Rat (7) ($[H^+] = 10^{-7.4}$)

$\log k_{Ac} + \log \frac{K_A + [H^+]}{[H^+]}$				
$= -1.260 + 0.308\Delta pK_A$	10	0.609	0.637	(27)
$= -1.102 + 0.821\pi$ $(\pm 0.573)(\pm 0.408)$	10	0.411	0.854	(28)
$= -0.732 + 1.271\pi - 0.228\Delta pK_A$	10	0.395	0.884	(29)

Rabbit (7) ($[H^+] = 10^{-7.4}$)

$\log k_{Ac} + \log \frac{K_A + [H^+]}{[H^+]}$				
$= -1.599 + 0.533\Delta pK_A$	10	0.606	0.821	(30)
$= -1.010 + 1.160\pi$ $(\pm 0.652)(\pm 0.465)$	10	0.468	0.898	(31)
$= -1.170 + 0.989\pi + 0.098\Delta pK_A$	10	0.493	0.901	(32)

Human (8) ($[H^+] = 10^{-7.4}$)

	n	s	r	
$\log k_{Ac} + \log \frac{K_A + [H^+]}{[H^+]}$				
$= -2.260 + 0.483\Delta pK_A$	10	0.627	0.783	(33)
$= -1.814 + 1.124\pi$ $(\pm 0.565)(\pm 0.402)$	10	0.405	0.916	(34)
$= -1.735 + 1.208\pi - 0.048\Delta pK_A$	10	0.431	0.916	(35)

The common feature of these correlations is that the ΔpK_A term plays no significant role. Since N-4-acetylation occurs by a nucleophilic attack of the lone pair of aromatic amino nitrogen, we could assume that the more the electron-donating property of the N-1-substituent, the faster the rate of acetylation at the N-4-position. However, separated from the benzene ring by the SO_2NH group, the electronic effect of the N-1-substituents may not be well transmitted to the N-4 position (*10*). Therefore, the lack of a significant role of the ΔpK_A term in these correlations shows that the most important factor governing the rate-determining step of the hepatic acetylation is the hydrophobicity of the drug.

The rate of renal excretion in the rat (*6*, *7*) is best correlated by Equations 38 and 44. The term correcting the effect of dissociation in the tubule is markedly significant in this case. Without this term, the correlation becomes much poorer (Equations 39–41). For rabbits (*7*), Equation 47 shows the best correlation. The coefficient of the ΔpK_A term in Equation 47 is negative, in contrast to that in Equations 38 and 44. Since the 95% confidence intervals of coefficients do not overlap, susceptibility of the tubular membrane to the electronic structure of permeating drugs probably differs between rats and rabbits.

Rat (7) ($[H^+] = 10^{-6.4}$)

	n	s	r	
$\log k_{Ex} - \log \frac{K_A + [H^+]}{[H^+]}$				
$= -0.715 - 0.229\Delta pK_A$	12	0.523	0.556	(36)
$= -0.732 - 0.706\pi$	12	0.324	0.858	(37)
$= -1.061 - 0.992\pi + 0.176\Delta pK_A$ $(\pm 0.602)(\pm 0.483)\ (\pm 0.242)$	12	0.299	0.893	(38)
$\log k_{Ex}$				
$= -1.254 + 0.037\Delta pK_A$	12	0.521	0.108	(39)
$= -0.913 - 0.160\pi$	12	0.509	0.234	(40)
$= -1.483 - 0.657\pi + 0.305\Delta pK_A$	12	0.455	0.565	(41)

Rat (6) ($[H^+] = 10^{-6.4}$)

	n	s	r	
$\log k_{Ex} - \log \frac{K_A + [H^+]}{[H^+]}$				
$= -1.476 - 0.119\Delta pK_A$	7	0.470	0.458	(42)
$= -1.386 - 0.427\pi$	7	0.332	0.779	(43)
$= -1.807 - 0.988\pi + 0.298\Delta pK_A$ $(\pm 0.563)(\pm 0.589)\ (\pm 0.279)$	7	0.207	0.937	(44)

Rabbit (7) ($[H^+] = 10^{-8.8}$)

$\log k_{Ex} - \log \frac{K_A + [H^+]}{[H^+]}$				
$= -0.680 - 0.628\Delta pK_A$	12	0.398	0.924	(45)
$= -1.639 - 1.193\pi$	12	0.493	0.880	(46)
$= -0.858 - 0.512\pi - 0.419\Delta pK_A$ $(\pm 0.693)(\pm 0.557)\ (\pm 0.279)$	12	0.345	0.949	(47)

Human (9) ($[H^+] = 10^{-6.0}$)

$\log k_{Ex} - \log \frac{K_A + [H^+]}{[H^+]}$				
$= -1.445 - 0.104\Delta pK_A$	9	0.377	0.395	(48)
$= -1.443 - 0.316\pi$	9	0.350	0.520	(49)
$= -1.514 - 0.432\pi + 0.058\Delta pK_A$	9	0.375	0.532	(50)
$= -1.284 - 0.292\pi'$	9	0.272	0.748	(51)
$= -1.514 - 0.487\pi' + 0.163\Delta pK_A$ $(\pm 0.572)(\pm 0.372)\ (\pm 0.252)$	9	0.247	0.830	(52)

For humans (9), the correlation is not as good as those found for rabbits and rats (Equations 48–50). This may be the result of the small variability of values on the left of the equation and of the difference in the ability of partition coefficient data to define permeability in the different test animals. If we use π' values derived from the partition coefficients with a $CHCl_3$–water system instead of those from an isobutyl alcohol–water system, Equation 52 results, which shows a reasonable correlation. This is the only example here where the π' values from a $CHCl_3$–water system show a better correlation than those from an isobutyl alcohol–water system. The susceptibility of human tubular reabsorption to the hydrophobicity of drugs might be different from those of the rat and rabbit; thus, the model with a $CHCl_3$–water system could simulate

the tubular membrane better than the model with an isobutyl alcohol–water system.

We omitted the data for acetosulfamine, sulfisoxazole, and sulfisomidine in deriving Equations 48–52 since an active tubular secretion mechanism has been postulated for their renal excretion (*9*) which would also suggest a functional difference of human tubular membrane from those of rats and rabbits. In fact, the data for these three compounds fitted poorly a preliminary corerlation with the use of π'.

Table I gives the results of calculations of the acetylation rate constants based on Equations 22, 28, 31, and 34. Acetosulfamine and sulfadimethoxine are omitted in the correlations of 12 compounds for which the substituent parameters are known since preliminary calculations showed that the data for these two did not fit well in each case. Different routes for the metabolism of these two compounds might be operative. In fact, for sulfadimethoxine, glucuronide formation has been shown to occur to a considerable extent, and this can not be neglected as compared with the acetylation at liver in rats (*6*) and humans (*11*). We expected that acetosulfamine would behave as an acetyl donor as well as the acetyl acceptor. Thus, the N-4-acetyl group of N-4-acetylacetosulfamine might result not only from acetyl-CoA but also from the N-1-acetyl group of another acetosulfamine molecule.

Table II shows renal excretion data. Calculated values were obtained from Equations 38, 44, 47, and 52. Here, the rate constants of acetosulfamine and sulfadimethoxine are as well correlated as those of the other compounds for rats and rabbits. This could be expected since the k_{Ex} value is directly determined by the proportion of the integrated amount of non-metabolized drug in the total urinary excreted materials whereas the k_{Ac} value is derived by assuming that metabolites, other than N-4-acetyl derivatives, are negligible in the urine. For humans, the rate constant of sulfadimethoxine is well correlated while that of acetosulfamine is not. The latter may be excreted by a different mechanism as mentioned.

There could be alternate models than those used in the above discussions of the elimination processes. For instance, a model of acetylation, where the binding equilibria of drugs with serum protein and liver tissues are slow reactions and compete with each other for drugs, would result in a different expression than Equation 14. However, analyses of the rates of acetylation and excretion for various test organisms, which are physicochemically as well as statistically reasonable, are only possible with the present models. The propriety of using these models would be further proved by the following discussions on the time of duration.

The most convenient parameter which expresses the time of duration is the value $t_{1/2}$, which is inversely proportional to the sum of the k_{Ac}

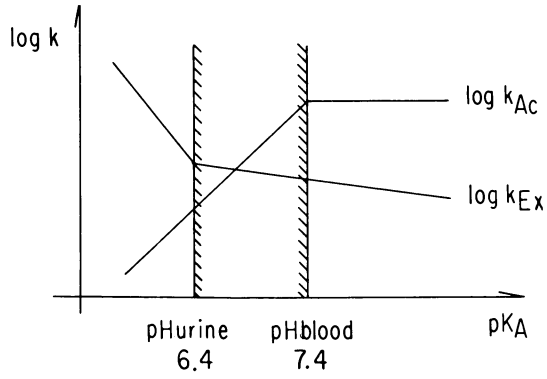


Figure 4. Dependence of log k *on* pK_A *(rat)*

and k_{Ex}. Although the values of log k_{Ac} and log k_{Ex} can be individually correlated by substituent constants, it would be mathematically as well as physicochemically difficult to analyze similarly the value of log (k_{Ac} + k_{Ex}). We did not try to analyze $t_{1/2}$, log $t_{1/2}$, or log (k_{Ac} + k_{Ex}) directly by substituent constants. Instead, we checked the physicochemical conditions which make both the log k_{Ac} and log k_{Ex} values as small as possible, simultaneously, so that the value of $t_{1/2}$ would be maximum.

By taking the partial differentials of Equations 28 and 38, we understand how the values of log k_{Ac} and log k_{Ex} for the rat vary according to the variations of the pK_A and π values of the N-1-substituent. Equations 53 and 54 are the partial differentials with respect to pK_A, which indicate that plots of both the log k values against pK_A consist of two phases as shown in Figure 4. The log k_{Ac} value decreases with a decline in the pK_A of drug beyond the value of blood pH while the log k_{Ex} value increases with a decreasing pK_A value beyond the urinary pH value. Since the urinary pH for the rat is about 6.4 ± 0.2 (which is smaller than that for blood which is *ca.* 7.4), we would expect that when the pK_A of a drug is located between these two pH values, both log k values of the drug would be as small as possible simultaneously.

$$\frac{\partial \log k_{Ac}}{\partial pK_A} = \frac{K_A}{K_A + [H^+]_{blood}} \tag{53}$$

$$\frac{\partial \log k_{Ex}}{\partial pK_A} = -\frac{K_A}{K_A + [H^+]_{urine}} - 0.176 \tag{54}$$

Comparing the partial differentials with respect to π of Equations 28 and 38 (Equations 55 and 56), we see that the dependence of both log k values on π is almost opposite. The larger the π value, the faster the rate of acetylation and the slower that of excretion. Thus, we could

not expect a hydrophobic property which would make both the log k values as small as possible. If, however, the small difference in the absolute values between partial differentials could be given a meaning, the larger the π value, the larger would be the time of duration since the susceptibility of tubular reabsorption to the hydrophobicity is larger than that of hepatic acetylation.

$$\frac{\partial \log k_{Ac}}{\partial \pi} = 0.821 \tag{55}$$

$$\frac{\partial \log k_{Ex}}{\partial \pi} = -0.992 \tag{56}$$

Figure 5 shows the $t_{1/2}$ values for the rat, obtained for 12 sulfonamide drugs (7), plotted against pK_A. This figure shows that the above discussions are qualitatively correct. The pK_A values of drugs which have

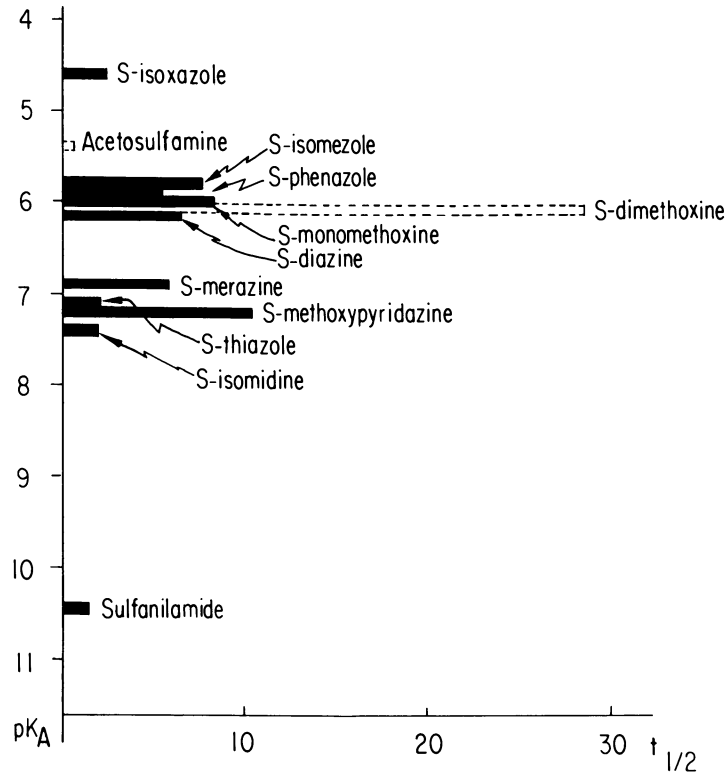


Figure 5. pK_A and time of duration (rat). For acetosulfamine and sulfadimethoxine where different routes of hepatic metabolism are suggested, the $t_{1/2}$ values are expressed by dotted columns.

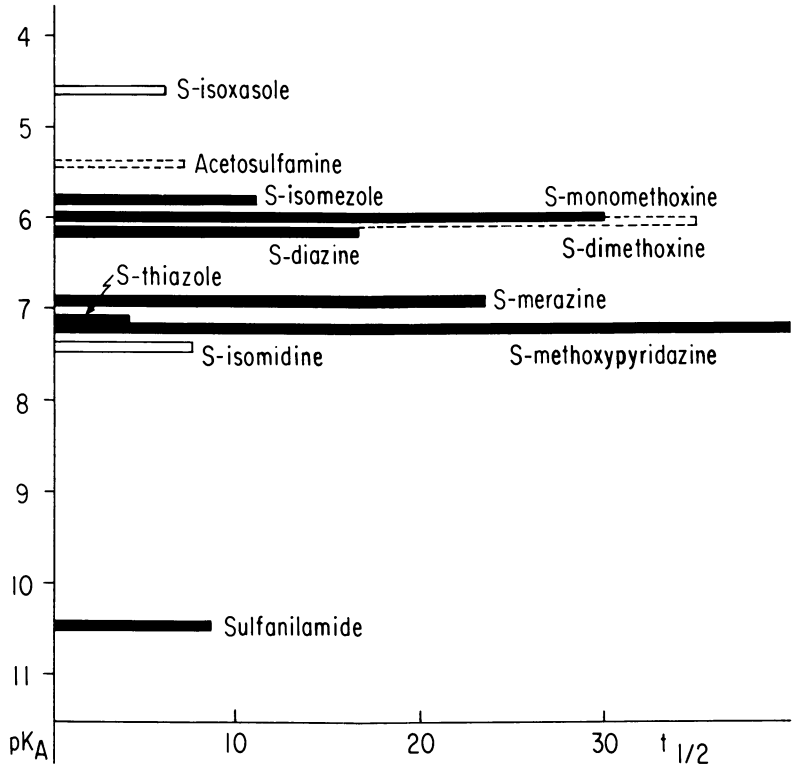


Figure 6. pK_A and time of duration (human). For acetosulfamine, sulfisoxazole, sulfisomidine, and sulfadimethoxine, where different mechanisms of renal excretion and hepatic metabolism are suggested by others, the $t_{1/2}$ values are expressed by blank and dotted columns; see *text.*

longer times of duration are mostly located in the range predicted above —*i.e.*, between the urinary and blood pH values, $pK_A = 6 \sim 7.4$. The times of duration of sulfathiazole and sulfisomidine which are quite short for their optimum pK_A values can be understood by their small π values. Since urinary and blood pH values are similar to those for the rat, the above discussion could also be applied to the time of duration for humans. Figure 6 shows that what we expect is also correct for the $t_{1/2}$ data obtained by Rieder and others (*4, 12*).

For the rabbit, the situation is different. Partial differentials of Equations 31 and 47 with respect to pK_A and π are given as Equations 57–60. According to Equations 57 and 58, we can draw biphasic plots

$$\frac{\partial \log k_{Ac}}{\partial pK_A} = \frac{K_A}{K_A + [H^+]_{blood}} \tag{57}$$

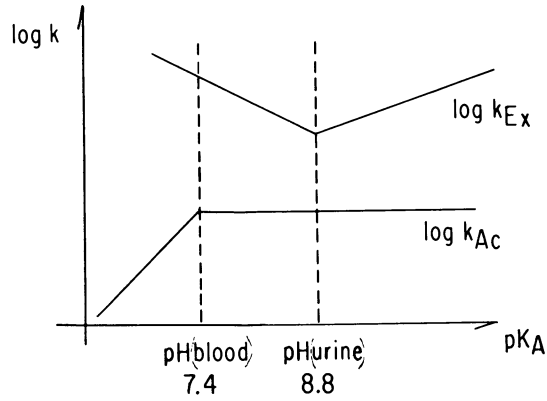


Figure 7. Dependence of log k *on* pK_A *(rabbit)*

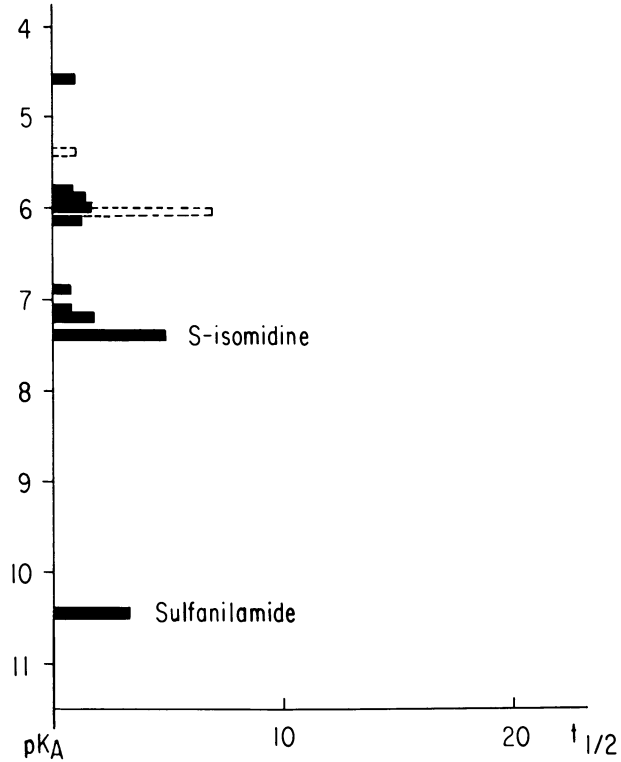


Figure 8. pK_A *and time of duration (rabbit). For acetosulfamine and sulfadimethoxine where different mechanisms of hepatic metabolism are suggested by others, the* $t_{1/2}$ *values are expressed by dotted columns.*

$$\frac{\partial \log k_{Ex}}{\partial pK_A} = -\frac{K_A}{K_A + [H^+]_{urine}} + 0.419 \quad (58)$$

for log k *vs.* pK_A, as shown in Figure 7. Here, the urinary pH, 8.8, is higher than that of the blood, 7.4. Hence, we could not expect any optimum range for the pK_A. Comparison of Equation 59 with 60 indicates

$$\frac{\partial \log k_{Ac}}{\partial \pi} = 1.160 \quad (59)$$

$$\frac{\partial \log k_{Ex}}{\partial \pi} = -0.512 \quad (60)$$

that hepatic acetylation is much more susceptible to hydrophobicity of the drug than is tubular reabsorption in rabbits. Therefore, the smaller the π value of the N-1-substituent, the larger the $t_{1/2}$ value would become. The reason why sulfanilamide and sulfisomidine, with π values which are among the smallest, have large $t_{1/2}$ values (*7*) (Figure 8) can be understood on this basis.

Analyses using free energy related electronic and hydrophobic substituent constants are capable of illustrating the rate constants of hepatic acetylation and renal excretion as well as the time of duration of sulfonamide drugs in various test organisms. The effects of drug dissociation on binding equilibria with various macromolecules and on permeability through membranes do play important roles in elimination processes. Although we made drastic assumptions in deriving the rate constant expressions, the reasonable correlations obtained show that these models for the drug elimination processes are practically useful and that other recently postulated elimination processes—*i.e.*, tubular active secretion (*9, 13, 14*) and N-1-glucuronidation followed by hepato-biliary excretion (*15*), need not be considered (except for a few compounds) as far as compounds studied in this work are concerned. Thus, we expect that the procedures developed here will provide information on the physicochemical properties which are fundamental in our search for new long-acting derivatives of the sulfonamide family as well as other dissociable drugs.

Acknowledgment

The author thanks Corwin Hansch for many suggestive discussions and for making the regression analyses. He is also grateful to Minoru Nakajima for his support of this work.

Literature Cited

1. Fujita, T., Hansch, C., *J. Med. Chem.* (1967) **10,** 991.
2. Fujita, T., *J. Med. Chem.* (1972) **15,** in press.

3. Krüger-Thiemer, E., Bünger, P., *Chemotherapia* (1965/1966) **10,** 61, 129.
4. Rieder, J., *Arzneimittel-Forsch.* (1963) **13,** 81.
5. Nelson, E., *J. Pharm. Sci.* (1961) **50,** 181.
6. Nogami, H., Hasegawa, A., Hanano, M., Imaoka, K., *Yakugaku Zasssi* (1968) **88,** 893.
7. Yamazaki, M., Aoki, M., Kamada, A., *Chem. Pharm. Bull.* (1968) **16,** 707, 721.
8. Kakemi, K., Arita, T., Koizumi, T., *Arch. Pract. Pharm.* (1965) **25,** 22.
9. Koizumi, T., Arita, T., Kakemi, K., *Chem. Pharm. Bull.* (1964) **12,** 428.
10. Yoshioka, M., Hamamoto, K., Kubota, T., *J. Chem. Soc., Japan* (1963) **84,** 412.
11. Bridges, J. W., Kibby, M. R., Walker, S. R., Williams, R. T., *Biochem. J.* (1969) **111,** 167.
12. Zbinden, G., ADVAN. CHEM. SER. (1964) **45,** 25.
13. Despopoulos, A., Callahan, P. X., *Am. J. Physiol.* (1962) **203,** 19.
14. Arita, T., Hori, R., Owada, E., Takahashi, K., *Chem. Pharm. Bull.* (1969) **17,** 2526.
15. Millburn, P., Smith, R. L., Williams, R. T., *Biochem. J.* (1967) **105,** 1283.
16. Yamazaki, M., Aoki, M., Kamada, A., Yata, N., *Arch. Pract. Pharm.* (1967) **27,** 37.

RECEIVED June 17, 1971.

7

Comparative *in vitro* and *in vivo* Structure–Activity Studies of Antiparasitic 2-Methyleneamino-5-nitrothiazoles

EBERHARD KUTTER, HANS MACHLEIDT, WOLFGANG REUTER, ROBERT SAUTER, and ALEXANDER WILDFEUER

Karl Thomae GmbH, Chemisch-pharmazeutische Fabrik, 795 Biberach an der Riss, West Germany

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

Eighteen nitrothiazoles have been prepared and tested for their trichomonicidal activity. Regression analysis and physicochemical parameters were used to correlate trichomonicidal in vivo *and* in vitro *activity against* T. foetus *with physicochemical properties of the nitrothiazole derivatives. A good correlation of* in vitro *trichomonicidal activity was obtained with log* P *and the oxidation-reduction potential* E_h *as the most important variables. This correlation, together with other findings, indicates that the 5-nitrothiazoles interfere with a normal metabolic redox process of the parasite. A more qualitative correlation was found between the oxidation-reduction potentials of different 5-nitroheterocycles and their antimicrobial activity against several types of microorganisms. Correlation of* in vivo *trichomonicidal activities in mice was much less significant. Experimental evidence is presented indicating that the kinetics of metabolism is the most important factor for* in vivo *activity. A study using optical isomers showed that metabolism proceeds highly stereospecifically.*

The main interest of scientists in the pharmaceutical industry is not only to produce biologically active compounds and to rationalize their effects from the physicochemical point of view but also effectively to develop drugs which might be used in human therapy. This paper gives a specific example of how this approach can increase the effectiveness and rationality of such drug development. Figure 1 shows, in a

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

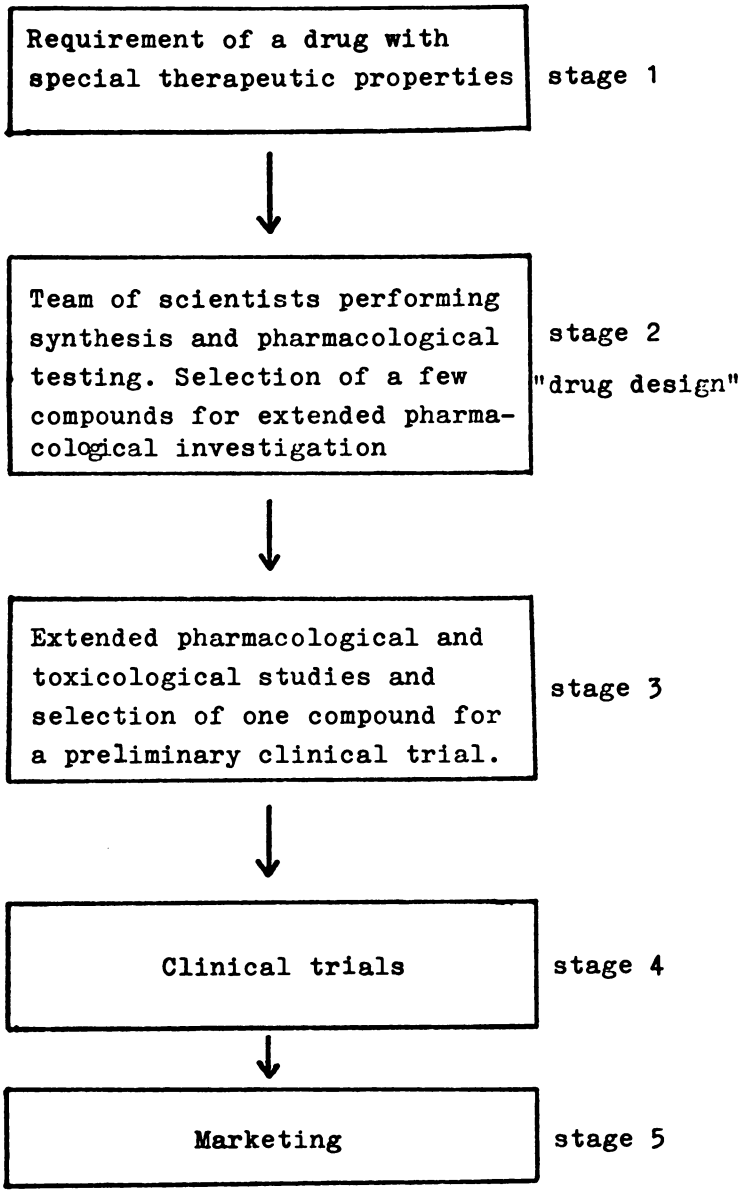


Figure 1. Several stages of drug development

simplified way, the stages of development a drug must go through before it is commercially useful.

As is immediately evident, the Hansch approach can only be helpful at the second stage of development. Many different compounds are pre-

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

pared and tested for their biological activity at this stage, and correlations between physicochemical properties and activities are only possible here. However, because of the tremendous amount of work which usually must be performed before a compound can be selected for extended investigation, stage 2 is one of the rate-limiting steps in the development of drugs to treat human diseases. Any approach which helps to accelerate development at this stage and to reduce the work necessary should lower overall research expenses.

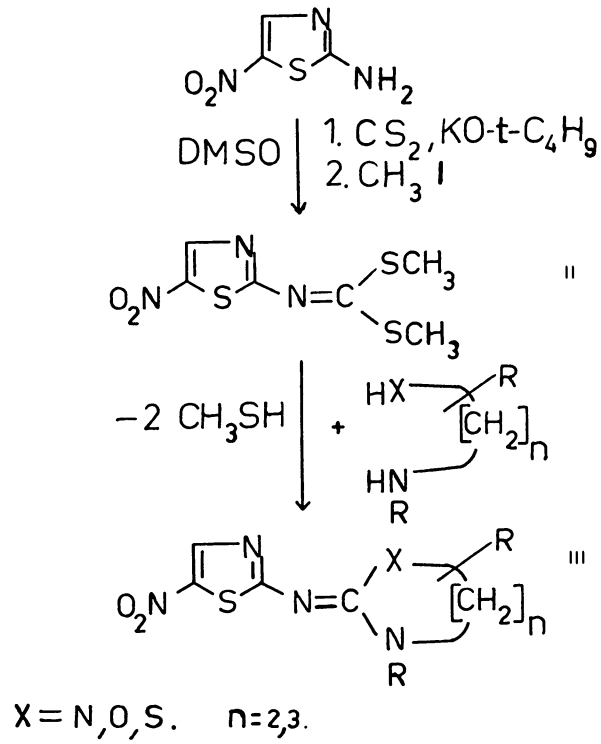


Figure 2. Synthesis of 2-methyleneamino-5-nitrothiazoles

Our example concerns a certain kind of antimicrobial agent and may illustrate how useful the Hansch approach actually can be in this second stage of drug development. We became interested in chemotherapeutic agents effective against trichomoniasis for several reasons. Trichomoniasis is a certain kind of venereal disease and is caused by the parasite *T. vaginalis*. In animals, infections can also be caused by another strain of parasite called *T. foetus*. We began synthesis in this field by using 2-amino-5-nitrothiazole as starting material (Figure 2).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

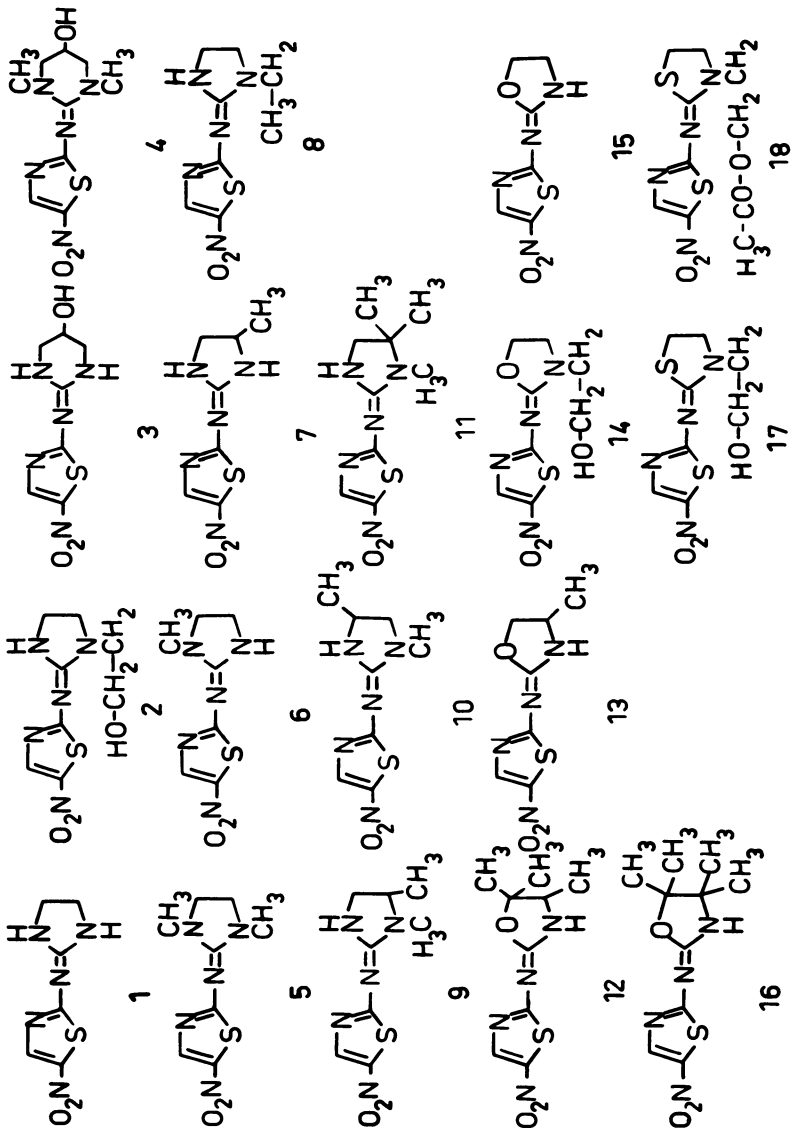


Figure 3. Chemical structures of 18 new nitrothiazole derivatives

2-Amino-5-nitrothiazole reacts with CS_2 and subsequently CH_3I to yield the dimethylthio derivative II. Ambifunctional nucleophilic amines react with this key compound II, with elimination of CH_3SH, and yield as the main product the 2-methyleneamino-5-nitrothiazoles, III. 2-Amino-5-nitrothiazole itself has a weak trichomonicidal effect. It was hoped that changes in the physicochemical properties which are brought about by the synthesis of the methyleneaminonitrothiazoles, III, would lead to more powerful drugs. Eighteen compounds which were synthesized by this route and tested *in vitro* and *in vivo* for their antitrichomonal activity are shown in Figure 3.

From Figure 3 it is obvious that the methyleneaminonitrothiazoles can be divided into three electronically different groups: those containing an amidine function, those containing an isourea function, and those containing an isothiourea function. For convenience we shall call them the N-type, the O-type, and the S-type compounds, respectively.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

Table I. Structure-Activity

Number [a]	*log* P [b]	E_h [c] (*volt*)	*MIC*- T. foetus [d] *μg/ml*	*log 1/C* [e] T. foetus (*obs.*)
1	1.38	−0.262	2.50	1.93
2	1.23	−0.262	2.50	2.01
3	0.74	−0.262	10.00	1.39
4	0.16	−0.262	100.00	0.43
5	0.92	−0.262	5.00	1.68
6	1.80	−0.262	2.50	1.96
7	1.82	−0.262	1.00	2.36
8	2.28	−0.262	1.66	2.16
9	2.02	−0.262	2.50	1.99
10	1.95	−0.262	1.66	2.16
11	2.35	−0.262	2.50	2.01
12	1.88	−0.232	1.00	2.41
13	1.34	−0.232	1.66	2.14
14	0.43	−0.232	5.00	1.71
15	0.57	−0.232	1.25	2.23
16	2.40	−0.232	1.25	2.33
17	1.28	−0.212	0.25	3.04
18	1.88	−0.212	0.125	3.40

[a] Coding according to Figure 3.

[b] Partition coefficients in the octanol–pH 7.4-phosphate-buffer system.

[c] Nitrothiazole oxidation-reduction potentials (volts) as calculated from their half-wave potentials, as determined using a Polarecord E 261 polarograph (Metrohm AG, Herisau, Switzerland) and a saturated Ag/AgCl reference electrode. Measurements were performed at 20°C and a drop time of 1 drop/2.8 sec. The compounds were dissolved in 1 ml dimethyl formamide and added to 24 ml of a borax–potassium biphosphate buffer of pH 7.3 [prepared according to J. M. Kolthoff, *J. Biol. Chem.* (1925) *63*, 135]. A pH of 7.4 resulted. Standard error of determination ±3 mv.

[d] Test organism: *Trichomonas foetus*. Minimal inhibition concentrations (μg/ml) were determined microscopically after 48 hr-incubation at 37°C. The two-fold serial

The *in vitro* biological activity was determined using the "serial dilution method" according to a standardized test procedure. The minimal inhibition concentrations (MIC values listed in Table I) represent the lowest concentration of a particular nitrothiazole derivative necessary to produce total inhibition of trichomonade growth after a fixed time interval. The advantage of this test procedure compared with other methods and the usefulness of MIC values in structure-activity studies were recently demonstrated by Seydel in an elegant study on sulfa drugs (*1*).

In starting correlation studies in the *in vitro* series of activities we first can compare the MIC values (Table I) in a more qualitative sense. At a first glance the S-type compounds seem to be the most active in this series, followed by the O-type. The N-type compounds, on the other hand, appear to be the least active *in vitro* from this point of view; however, there are certain exceptions to this rule, and therefore only a quantitative correlation can reveal useful structure activity information.

Parameters for Nitrothiazoles

log 1/C[f] T. foetus (*pred.*)	Δ *log 1/C*	*ED* $_{50 \pm 17}$[g] T. foetus *mg/kg*	*Acute Toxicity*[h] *Mice p.o.* LD_{50} *g/kg*
1.98	0.05	>250	>4.0
1.90	0.11	75	5.64
1.46	0.07	75	—
0.62	0.19	>100	—
1.65	0.03	37.5	1.0
2.09	0.13	37.5	7.6
2.09	0.27	>100	>4.0
1.98	0.18	37.5	>4.0
2.07	0.08	50.0	>10.0
2.08	0.08	37.5	>10.0
1.94	0.07	100.0	>4.0
2.70	0.29	37.5	>25.8
2.59	0.45	75.0	1.26
1.67	0.04	100.0	2.89
1.87	0.36	75.0	1.94
2.53	0.20	25.0	0.05
2.96	0.08	37.5	1.00
3.12	0.28	37.5	2.18

dilution method in fluid thioglycollate medium (with 10% horse serum) was used, and 20,000 protozoa were inoculated per ml medium. The compounds were dissolved in Titrisol buffer pH 7 (Merck Co., Darmstadt, West Germany) and tested at the following concentrations (μg/ml): 10, 5, 2.5, 1.66, 1.25, 1.0, 0.5, 0.25.

[e] *In vitro* activities against *T. foetus*. *C* is represented in μg/ml × mw units.

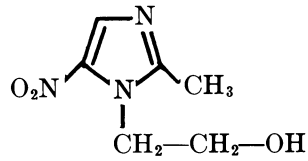
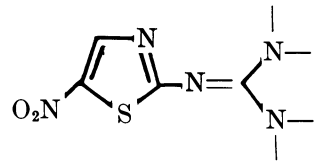
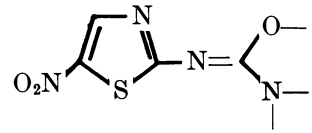
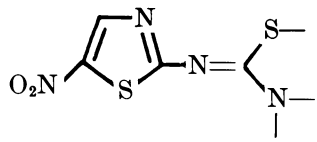
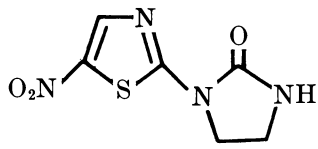
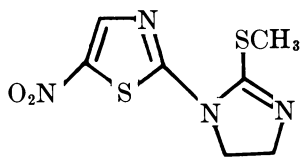
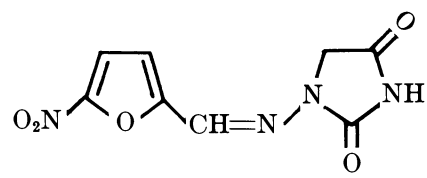
[f] Calculated using Equation 5.

[g] $ED_{50 \pm 17} = 50 \pm 17\%$ survival rate of mice (NMRI, 18–20 g, female) infected with 1–2 × 10^6 trichomonades. Dosage schedule: Orally twice daily for 3 days. The first treatment followed 2 hr after infection. The untreated control mice died within 5 days.

[h] 14 days LD_{50} in g/kg. Mice, NMRI, female and male, 18–20 g. The substances were suspended in tylose and given by gavage.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table II. Antibacterial and Antiprotozoal Activity of Several

	E_h (*mv*)
	−354
	−262
	−232
	−212
	−211
	−186
	−47

5-Nitroheterocycles Compared with Their Oxidation Potentials

E. coli *ATCC 9637* in vitro	Sc. Aronson in vitro	T. foetus T. vaginalis in vitro
inactive	inactive	moderate
inactive	inactive	moderate
inactive	inactive	good
inactive	inactive	high
moderate	moderate	high
weak	inactive	high
moderate	good	weak

	Bacteria	*Trichomonades*
inactive	>80 μg/ml	>10 μg/ml
weak activity	80 μg/ml	10 μg/ml
moderate activity	25–5 μg/ml	10–1.0 μg/ml
good activity	5–1 μg/ml	1.0–0.2 μg/ml
high activity	<1 μg/ml	<0.2 μg/ml

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

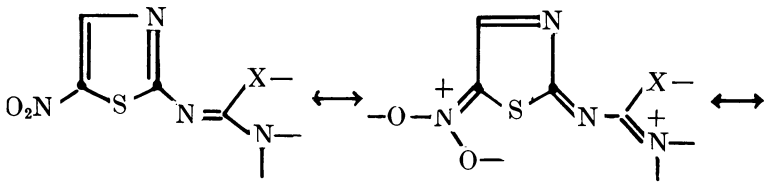
Before starting quantitative correlation studies, one must find physicochemical parameters which may be meaningful in interpreting the variance in the biological activities. In this special case we are faced with the tricky problem of intercorrelation among several electronic parameters. Any physicochemical property in which oxygen takes an intermediate position between sulfur and nitrogen should be suitable for correlating the electron dependent differences in activity. Fortunately, other experimental findings help to overcome this difficulty. From an elegant study with an antibacterial 5-nitrofuran derivative, which is chemically quite similar to the 5-nitrothiazoles, Cramer (*2*) concluded that this nitrofuran derivative interferes with a normal metabolic process of bacteria and does so by virtue of its reducibility. This hypothesis is supported by the studies of Hirano *et al.* (*3*) and Powers and Mertes (*4*) on the mechanism of action of other antibacterial nitroheterocycles. The physical significance of the oxidation-reduction potential in the mode of action of the 5-nitroheterocycles is also stressed by the more qualitative correlation we found between the oxidation potentials of different 5-nitroheterocycles and their antimicrobial activity against several types of microorganisms.

In Table II the activities of some 5-nitroheterocycles against bacteria and protozoa are compared with their oxidation potentials. Interesting conclusions can be drawn from this table in regard to the influence of the oxidation-reduction potential of the drugs on their specificity of action against several types of microorganisms. The 5-nitroimidazole derivative and our N-type derivatives have the lowest oxidation potential of this series. These compounds possess only trichomonicidal activity. However, with an increase in the oxidation potential not only does the *in vitro* trichomonicidal activity increase but selectivity decreases, and more and more types of microorganisms are affected. Thus, in this series the compound with the highest oxidation potential, the 5-nitrofuran derivative, is active against all types of organisms in this table. Organisms with a higher degree of evolutionary development seem to be influenced even by nitroheterocycles with a relatively low oxidation potential. On the other hand, the "nitro reductase" (*2*) of the bacteria seem to possess a low reduction potential; therefore one needs a nitroheterocycle with a high oxidation potential to inhibit bacterial metabolism.

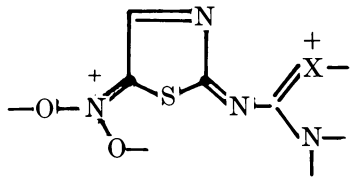
Structure IVc makes different contributions, depending on the nature of the heteroatom X. Therefore, the oxidation potential of these drugs is predominantly determined by the heteroatom X in structure IV; ring size and alkyl substitution are secondary effects. The oxidation potentials of the compounds 2, 14, and 17 were measured since they were soluble.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.



IV a IV b



IV c

Their values were chosen as a first approximation to represent the redox potentials of these three kinds of compounds in a Hansch-type multiple regression analysis study.

$$\log 1/C = [21.23(\pm 13.28)]\, E_h + 7.34(\pm 3.30) \quad (1)$$
$$n = 18 \quad s = 0.486 \quad r^2 = 0.418 \quad r = 0.646$$

$$\log 1/C = [0.494(\pm 0.40)] \log P + 1.35(\pm 0.65) \quad (2)$$
$$n = 18 \quad s = 0.534 \quad r^2 = 0.299 \quad r = 0.547$$

$$\log 1/C = -[0.64(\pm 0.57)]\, (\log P)^2 + [2.20(\pm 1.57)] \log P + 0.50(\pm 0.95) \quad (3)$$
$$n = 18 \quad s = 0.470 \quad r^2 = 0.491 \quad r = 0.700$$

$$\log 1/C = [0.524(\pm 0.248)] \log P + [22.14(\pm 9.00)]\, E_h + 6.80(\pm 2.25) \quad (4)$$
$$n = 18 \quad s = 0.327 \quad r^2 = 0.753 \quad r = 0.868$$

$$\log 1/C = -[0.54(\pm 0.28)]\, (\log P)^2 + [1.96(\pm 0.77)] \log P + [20.81(\pm 6.37)]\, E_h + 5.76(\pm 1.68) \quad (5)$$
$$n = 18 \quad s = 0.229 \quad r^2 = 0.887 \quad r = 0.942$$

$$\log P_o = 1.83(1.6\text{—}2.4)$$

In Equation 1, C is the molar MIC *in vitro* value, and E_h represents the oxidation-reduction potentials of the drugs. The figures in parenthesis are the 95% confidence limits. Equation 1 is statistically significant. Of the variance in the *in vitro* log $1/C$ values 42% can be explained by a linear regression with E_h. This supports the hypothesis that the 5-nitrothiazoles affect an important enzymatic reduction process of the tricho-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

monades. However, 58% of the variance in the biological result remains unexplained by Equation 1. This indicates that the oxidation-reduction potential is not the only physicochemical parameter which determines the *in vitro* activities against *T. foetus*.

Reductases are usually part of multienzyme complexes which are lipophilic constituents of a cell. Compounds which act on these lipophilic structures should also show appreciable lipophilic properties in order to be bound by them. According to Hansch (*5*, *6*) the lipophilic character of compounds can be represented by their partition coefficients in the octanol-water system. Therefore we have measured the partition coefficients for our nitrothiazoles in an octanol–pH 7.4 buffer and included them in our regression study. A comparison between the two single-variable equations, 1 and 2, shows that the electronic and the lipophilic parameters are approximately of equal importance. Although neither correlation is highly significant, the linear combination of the two gives a good result in Equation 4. However, the addition of a $(\log P)^2$ term in Equation 5 is a significant improvement and results in an excellent correlation that explains approximately 90% of the variance in the biological data. If we take into account the inaccuracies of the biological data and the simplifying assumptions as far as E_h is concerned, the quality of this correlation is as high as can be expected in this biological system.

Equation 4 implies a parabolic relationship between activity and log *P*—a phenomenon often observed in quantitative structure–activity studies (*7*, *8*). Polar drugs can penetrate cell walls only slowly (*9*). With an increase in the lipophilic character of the drugs an increase in penetration (and therefore in activity) is observed. However, activity cannot continuously increase linearly with the lipophilic character of the molecules. The biological material also is heterogeneous as far as its lipophilic character is concerned. Membranes, organelles, or multienzyme complexes, for example, are very lipophilic cell components, and therefore very lipophilic drugs are preferentially bound. On the other hand, enzymes of the plasma, nucleic acids, or proteins of the culture medium have a lower degree of lipophilic character, and therefore compounds with intermediate lipophilic properties are preferentially bound by the latter components. This explains why an optimal range of lipophilicity exists for drug molecules acting on a certain receptor compartment. From Equation 5 it can be concluded that compounds with partition coefficients of 1.83, with the 95% confidence limits ranging from 1.6 to 2.4, should have especially favorable trichomonicidal activities. In Table I the observed *in vitro* trichomonicidal activities and those predicted from Equation 5 are compared. Generally a good agreement between measured and calculated biological activities is observed.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

In a development stage, when we had data for only 8–10 points meaningful correlations were still possible. Thus, equations analogous to Equation 5 have been effective guides to further synthesis of nitrothiazoles possessing high activity *in vitro*.

To be able to compare *in vitro* and *in vivo* correlations we also tested all 18 nitrothiazoles *in vivo*. Mice were infected intraperitoneally with a lethal amount of trichomonades and then treated with different dosages of a nitrothiazole derivative.

As the *in vivo* data in Table I show, great differences compared with *in vitro* activities are immediately discernible. Thus, there are now highly active compounds in all three groups. On the other hand, some of the compounds with good *in vitro* activities have proved to be inactive *in vivo*. Therefore, it was interesting to check whether those parameters which are relevant *in vitro* are also useful in explaining the variance in the *in vivo* data.

$$\log (1/C) = -[0.132(\pm 0.29)] (\log P)^2 + [0.562(\pm 0.86)] \log P + [3.37(\pm 5.27)] E_h + 1.02(\pm 1.40) \quad (6)$$
$$n = 15 \quad s = 0.173 \quad r^2 = 0.443 \quad r = 0.665$$

Naturally the three compounds which were inactive *in vivo* could not be included in the *in vivo* correlation. Equation 6 is only significant at the 90% confidence level. In the two equations, coefficients of the same parameters and intercepts have identical signs. However, in Equation 6, coefficients generally are lower, and confidence limits are larger than those in Equation 5. Moreover, Equation 6 can only explain 44% of the variance in the *in vivo* data while Equation 5 explains 89% of the variance in the *in vitro* data. Therefore it is no surprise that Equation 6, unlike Equation 5, cannot be used for predictions. Of course, a high oxidation potential and a certain degree of lipophilicity are also important factors for high *in vivo* activity. However, the *in vivo* system is much more complicated than the *in vitro* system, and therefore additional parameters come into play. Not only do the nitrothiazoles affect the parasites *in vivo* but the drugs themselves are affected by the host. The natural defense mechanisms can only overcome the infection when an active blood and tissue level of a particular drug can be maintained for a sufficiently long time. Therefore, the pharmacokinetic properties of the nitrothiazoles are decisive as concerns the *in vivo* activity of the *in vitro* active derivatives.

We wished to discover which pharmacokinetic property is of special importance for nitrothiazole *in vivo* action. In a separate experiment it was shown that particle size had no decisive influence on *in vivo* activity. Increasing experimental evidence indicates that absorption and distribu-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

tion of drugs which follow a passive penetration mechanism are largely governed by their lipophilic properties (*10, 11*). In addition, some authors have shown that metabolism and elimination are also a function of the lipophilicity of the drugs (*12, 13*). As mentioned earlier, the lipophilic properties of the nitrothiazoles should be reflected by their log *P* values. Log *P* is already included as a parameter in the *in vivo* correlation, Equation 6. However, the poor quality of this correlation indicates that this parameter does not account for all of the pharmacokinetic properties of the nitrothiazoles.

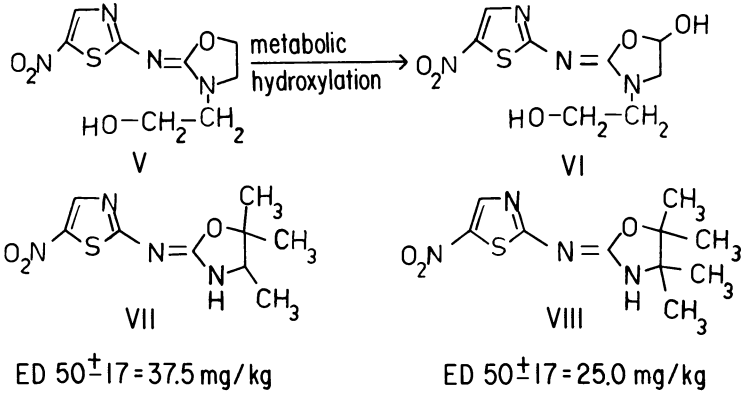


Figure 4. Metabolic behavior and activities of some O-type nitrothiazole derivatives

From the results of preliminary metabolic studies we concluded that the kinetics of metabolism of the nitrothiazoles cannot only be controlled by the lipophilicities of these drugs. This hypothesis was supported by two independent experimental findings. In Figure 4 structure VI represents a metabolite of an O-type derivative. V is subject to metabolic hydroxylation at a ring CH_2 group. The metabolite VI can be found in high concentrations in the urine of mice. This indicates that one of the main metabolic pathways of degradation of the O-type derivatives is a metabolic attack at the ring CH_2 groups. On the basis of this finding the high *in vivo* activity of the two highly methylated O-type derivatives VII and VIII, which is surprising in light of their moderate *in vitro* activity, can easily be explained by a blockade of an important metabolic degradation pathway for these drugs. Additional support for the hypothesis that steric factors are important in nitrothiazole metabolism comes from a study using optical isomers.

The two optical isomers shown in Figure 5 were synthesized and tested for their trichomonicidal activity *in vitro* and *in vivo*. Both isomers have the same oxidation potentials and identical partition coefficients.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

For high *in vitro* activity only these two parameters were significant. Therefore it is not surprising that the two optical isomers show identical *in vitro* trichomonicidal activities; however, there is a highly significant difference in their *in vivo* activities.

Penetration and distribution properties should show no difference for the two isomers since they both have identical partition coefficients. However, one could argue that the passive penetration process could be different for the two optical isomers because of a certain stereoselective preference of absorption for one of the isomers by proteins. This seems highly unlikely taking into account the insignificance of these effects (*14*). In addition, examples *in vivo* are known where no difference in penetration properties for optical isomers could be observed (*8*, *15*). On the other hand, if these effects contribute significantly to the penetration phenomenon, one would at least expect an indication of such an effect in the protein-rich *in vitro* culture medium. However, in comparing *in vitro* and *in vivo* data in Figure 5, nothing can be found to support this hypothesis.

Again, metabolism seems to be important in rationalizing the observed differences in the *in vitro* and *in vivo* data. If one assumes that the metabolic enzymes show some degree of stereospecificity, differences in the kinetics of metabolism for the two optical isomers would result. A kinetic study with mice on the amount of certain metabolites excreted by the kidney confirms this point of view. All nitrothiazoles of the O-type have an ultraviolet (UV) maximum at 375 mμ. Therefore all metabolites which still contain this chromophore, which is characteristic for the O-type compounds, can be collectively determined UV-spectroscopically. A comparative kinetic study of the two isomers over a 24-hour period reveals different metabolic behavior. This strongly supports the hypothesis that the kinetics of metabolism is an important factor for nitrothiazole *in vivo* activity. Moreover one can conclude that in this case metabolism proceeds highly stereospecifically.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

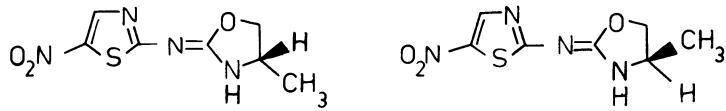


Figure 5. Structure-activity parameters of two optical isomers

S configuration	R configuration
$[\alpha]_D^{25} - 51.7°$ (c = 0.1 g/100 ml, ethanol)	$[\alpha]_D^{25} + 51.4°$ (c = 0.1 g/100 ml, ethanol)
MIC (T. foetus): 1.62[a] μg/ml	MIC (T. foetus): 1.88[a] μg/ml
ED_{50} (mice): 145 mg/kg (127–165)[b]	ED_{50} (mice): 61 mg/kg (46–81)[b]

[a] *Mean values, difference not significant at the 95% confidence level*
[b] *95% confidence limits (Lichtfield and Wilcoxon)*

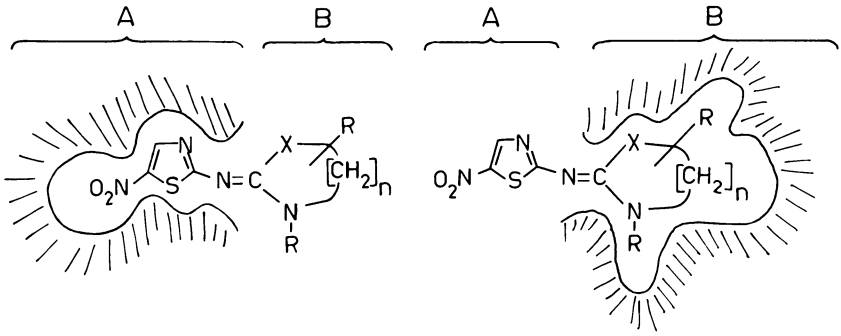
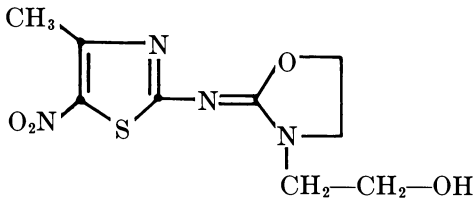


Figure 6. Simplified models for the steric requirements of the "nitroreductase" of the trichomonades (left) and of the metabolic enzymes receptor site (or sites) in mice (right)

In Figure 6, the complexity of the *in vivo* situation compared with *in vitro* becomes obvious. The *in vivo* data are the result of the interaction of the nitrothiazoles with at least two different kinds of receptors. As can be seen in Figure 6, the steric requirements for the two sites are different. The inactivity of compound IX suggests that part A of the



IX: inactive against *T.foetus* at 100 μg/ml

nitrothiazole molecule is attached stereospecifically to the nitroreductase of the trichomonades while part B can be visualized as the transport part of the nitrothiazoles III, which indirectly can also influence the electronic properties of part A. In the set of nitrothiazoles in Table I substituents were only changed in part B of the nitrothiazole moiety and the steric properties of part A were kept constant. Therefore, the stereospecific requirements of the interaction of part A with the nitroreductase are not reflected in Equations 5 and 6. On the other hand, as is obvious from the constitution of the metabolite VI and the study with the optical isomers, part B of the nitrothiazoles seems to be attached stereospecifically to the metabolic enzymes of the host. The Hansch model is developed under the simplifying assumption that there are critical steps at only one site of action which are decisive for the macroscopically observable biological result (*16*). On the basis of the above

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

considerations it seems that in some cases this model has to be refined to account for more complex situations.

Conclusions

The Hansch approach can be a useful and time-saving method which facilitates the complicated task of drug design. Taking into account the large number of new nitrothiazoles which could be conceived, the usefulness of the *in vitro* correlation as a predictive guide for further synthesis cannot be overestimated. On the other hand, one should not be too surprised if with a limited set of data points no significant *in vivo* correlation can be found. The complex nature of an *in vivo* result may often lead to a situation where no great difference exists between the amount of biological data at hand and the numbers of parameters necessary to explain the variance in the data. Of course, a good correlation cannot be expected under these circumstances. Although good *in vivo* correlations were found in several instances (*8, 17–19*), a major breakthrough in this field presupposes an increase in our knowledge as far as the correlations between the kinetics of metabolism and the chemical structures of drugs in general are concerned. This indicates the importance of suitable *in vitro* test systems. Physicochemical parameters which have proved to be important for high *in vitro* activity can then be used as a further guide for the synthesis of compounds highly active *in vivo*.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch007

Literature Cited

1. Seydel, J. K., *J. Med. Chem.* (1971) **14,** 724.
2. Cramer, D. L., *J. Bacteriol.* (1947) **54,** 119.
3. Hirano, K., Yoshina, S., Okamaru, K., Suzuka, I., *Bull. Chem. Soc. Jap.* (1967) **40,** 2229.
4. Powers, L. J., Mertes, M. P., *J. Med. Chem.* (1970) **13,** 1102.
5. Fujita, T., Iwasa, J., Hansch, C., *J. Amer. Chem. Soc.* (1964) **86,** 175.
6. Leo, A., Hansch, C., Church, C., *J. Med. Chem.* (1969) **12,** 766.
7. Lien, E. J., Hansch, C., Anderson, S. M., *J. Med. Chem.* (1968) **11,** 430.
8. Kutter, E., Herz, A., Teschemacher, H. J., Hess, R., *J. Med. Chem.* (1970) **13,** 801.
9. Garrett, E. R., Chemburkar, P. B., *J. Pharm. Sci.* (1968) **57,** 1401.
10. Seydel, J. K., *Proc. 6th Int. Congr. Chemotherapy* (Tokyo) 1969, S. 881.
11. Raaflaub, J., *Experientia 26-Fasc.* (1970) **5,** 457.
12. Brodie, B. B., Gilette, J. R., La Du, B. N., *Ann. Rev. Biochem.* (1958) **27,** 427.
13. Hansch, C., Lien, E. J., Helmer, F., *Arch. Biochem. Biophys.* (1968) **128,** 319.
14. Bradley, W., Easty, G. C., *J. Chem. Soc. (London)* **1953,** 1519.
15. Wahlström, G., Buch, H., Buzello, W., *Acta Pharmacol. Toxicol.* (1970) **28,** 493.
16. Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., Streich, M., *J. Am. Chem. Soc.* (1963) **85,** 2817.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

17. Hansch, C., Steward, A. R., Anderson, S. M., Bently, D., *J. Med. Chem.* (1968) **11**, 1.
18. Hansch, C., Steward, A. R., Iwasa, J., *Mol. Pharmacol.* (1965) **1**, 87.
19. Hansch, C., Lien, E. J., Helmer, F., *Arch. Biochem. Biophys.* (1968) **128**, 319.

RECEIVED December 27, 1971.

8

Comparison of the Hansch and Free-Wilson Approaches to Structure-Activity Correlation

PAUL N. CRAIG

Craig Chemical Consulting Services, Inc., Ambler, Pa. 19002

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

The basic principles on which the Hansch multiple parameter method for structure–activity correlation depends are described. These are compared with the basic features of the Free-Wilson method for assigning additivity constants to structural features of related compounds. An example is given for which the two methods of analysis have led to similar structure–activity relationships. Factors which determine the particular method to use in a new situation are discussed. The Free-Wilson method is presented in considerable operational detail with special emphasis on the detection and avoidance of situations which lead to singularity problems in solution of the matrix. Favorable analyses, which result in additivity constants that can be correlated with known physical constants, may lead to predictions for new compounds not covered in the original matrix.

The two methods of structure–activity correlation which have received the most application in the past decade are the Hansch multiple parameter method, or the so-called extrathermodynamic approach, and the Free-Wilson, or additive model. The basic differences and similarities of these methods are discussed in this presentation.

The Hansch Multiple Parameter Method

Since this entire book has been structured around this method, it will not be discussed in great detail, but the highlights of the method will be summarized. The reader is referred to an excellent review by Hansch (*1*), which both describes the method and places it in a proper historical perspective.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

The multiple parameter approach tests for simple mathematical equations which can relate the biological activities of a series of closely related compounds to one or more physical parameters, which may be measured or calculated for these compounds. The parameters may be used singly or together, or in combination as linear and squared terms which can show parabolic relationships. Many possible combinations of parameters can be considered, and multiple regression analysis is employed to obtain statistical parameters for all combinations of parameters studied. The biological data are assumed to be more variable and less accurately determined than the physical parameters. Hence the biological data are assigned the role of the dependent variable, and the physical parameters are considered as independent variables in the regression.

The statistical parameters which result from the regression enable one to reject those relationships which are not statistically significant and to choose from those equations, which do pass standard statistical tests for significance, the ones which best explain the observed biological data. Of course, common sense must still play a role in the evaluation of meaningful equations, but the statistical parameters can be powerful guides, especially in rejecting or questioning the validity of relationships expressed by the mathematical equations. Typical equations which often are found to relate biological activities with physical parameters include:

$$\text{Log } \frac{1}{C} = a + b\pi \tag{1}$$

$$\text{Log } \frac{1}{C} = a + b\pi + c\sigma \tag{2}$$

$$\text{Log } \frac{1}{C} = a + b\pi - c\pi^2 + d\sigma \tag{3}$$

$$\text{Log } \frac{1}{C} = a + b\pi - c\pi^2 + d\sigma + eE_s \tag{4}$$

The symbols π, σ, and E_s refer to the substituent constants for partition, polar, and steric factors (*1*).

In practice, what is sought is usually the simplest equation which is not improved by addition of further terms. By "improvement" is meant a statistically significant reduction in the overall variance. When such an equation is obtained, substitution of the physical parameters for substituents not yet studied can be made, and the equation leads to a prediction of the biological activity of the unprepared compound.

Additional structure–activity information can be obtained when a parabolic relationship in the partition term is observed, provided the sign

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

of the coefficient for the squared term is negative. An optimal value for the partition factor can be obtained by differentiating the equation with respect to π or P; this results in the optimal P value (log $P°$). This can be a valuable guide in the design of new molecules which may differ considerably in structure from those studied in the regression analysis. If a positive coefficient is obtained for the square of partition term, the equation must be rejected as it implies that any change in partitioning will result in an increased biological activity; such a minimum has never been encountered. Of course this applies to the use of log $1/C$ as the method for quantitating the biological activities of the compounds under study.

The most common error in application of this method lies in a lack of appreciation of the minimum statistical requirements involved. Thus, one needs to have about five well-chosen compounds for every variable term in a Hansch analysis in order to feel confident about the results. For example, an equation such as Equation 2 above should be derived from 10 or more compounds, and one such as Equation 3, from 15 or more examples. A smaller number of examples per term may lead to useful results, but one cannot often support these results by statistics. A frequent abuse is seen when a large number of variable terms are used in a complex equation (four or more terms) which was derived from only 10 or 12 examples. The statistician would prefer to have 15 to 20 more compounds than the degrees of freedom in the resulting equation; not often is this luxury met.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

The Free-Wilson Additivity Model

Unlike the Hansch approach, in this model no assumptions are made concerning physical parameters which may play a role in determining the biological activity. Instead, a series of *de novo* substituent constants is obtained using only the experimentally obtained biological test data and the following basic assumption: every time a particular substituent group appears at the same place in the molecule, it is assumed that it will play a constant role towards determining the over-all biological activity of the molecule (*2*). It may contribute to, or detract from, the over-all biological activity, but it must always play the same role.

This basic assumption is checked by means of the statistical parameters which result from solution of the matrix, which expresses the assumption stated above in the following equation for each compound:

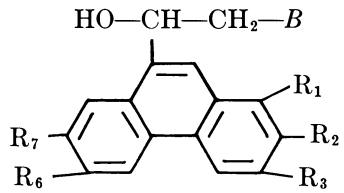
$$\text{Biological Activity} = \mu + \Sigma\, G_i X_i$$

where μ is the average biological activity, and G_iX_i represents the activity contribution for the i^{th} group at the i^{th} position. In structuring the matrix,

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

X becomes 0 or 1, indicating the absence or presence of a particular group at position X. The matrix thus represents a series of equations in multiple unknowns, one equation for each compound. Its solution gives the values for the *de novo* substituent constants for every substituent at each position ($= G_iX_i$). (A more complete discussion is given below.)

If the statistical parameters obtained upon solution of the matrix indicate that the additivity assumption is valid, the *de novo* constants can then be used to predict the activity of (a) those compounds used in derivation of the constants and (b) all possible combinations of the various groups at each position. This is not much of a saving when one has only two or three positions of a molecule which can be substituted, but in more complex situations this can be a powerful tool. An example is given below, where six different positions of the phenanthrene ring were substituted with three, three, six, three, six, and three substituents, respectively:



B = 3 groups	R_3 = 6 groups
R_1 = 3 groups	R_6 = 6 groups
R_2 = 3 groups	R_7 = 3 groups

This represents $3 \times 3 \times 3 \times 6 \times 6 \times 3 = 2916$ possible compounds. A good Free-Wilson analysis was obtained from only 42 of the possible analogs; the preparation of these 42 compounds enables one to predict with a fair assurance the approximate antimalarial activity to be expected for almost 2900 unprepared analogs.

The minimum number of compounds required for a Free-Wilson type of analysis will vary depending upon the number of positions substituted and the number of substituents at each position. The formula for the absolute minimum required to permit a solution of the complex set of equations in multiple unknowns is $N = 1 + (A - 1) + (B - 1) + (C - 1) + \ldots$, where A, B, C are the number of substituents at each position. This minimum should be exceeded by 10–20 compounds although useful results can sometimes result from as few as five compounds in excess of the minimum.

The fundamental point which differentiates between the two methods is the following: the Hansch method seeks for correlations between

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

variable biological activities and variable physical parameters whereas the Free-Wilson method uses only the biological activities as variable terms, along with exact information as to the presence or absence of each substituent group. Therefore, the *de novo* substituent constants which result embody all factors, known or unknown, that play roles in determining the biological activity of the particular compounds under study.

It is obvious that there are cases where groups will interact with each other and hence cannot have only additive effects. Fujita and Ban (*3*) have reported a successful attempt to include possible interactions in a Free-Wilson type of analysis. In a study of a series of substrates for dopamine-β-hydoxylase, a series of dopamine analogs was studied. In addition to the usual Free-Wilson matrix a term was added to allow for the interaction which might be possible from having two hydroxyl groups placed ortho to each other or from a hydroxyl group ortho to a methoxyl group. These terms were added to the regular matrix as a hypothetical new position; the statistical parameters were then compared with those for the conventional run. The addition of the term expressing the ortho relationship of the methoxyl and hydroxyl groups did give a significant improvement to the correlation whereas the term expressing the ortho relationship of two hydroxyl groups did not improve the regular correlation. The value for the interaction term had a negative coefficient, showing that this interaction resulted in a reduction of the biological activity.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

Overlaps Between the Two Methods

Cammarata compared the Free-Wilson constants derived for a series of tetracycline analogs with some physical constants, and found a relationship which involved two parameters (*4*). In as yet unpublished work on antimalarial compounds I have found good linear relationships to exist between certain Free-Wilson substituent constants (S.C.) and Hansch's "pi" or Hammett's "sigma" constants for the same substituents. These relationships are shown in Table I.

Table I. Relationships among Parameters

Group	R_3 *S.C.*	*Pi*	*Sigma para)*	R_6 *S.C.*
CF_3	0.332	1.16	0.54	0.476
Br	0.223	0.86	0.23	0.388
Cl	0.0688	0.71	0.23	−0.118
H	−0.257	0	0	−0.431
F	−0.265	0.14	0.06	−0.159
OCH_3	–	−0.02	−0.27	−0.510

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

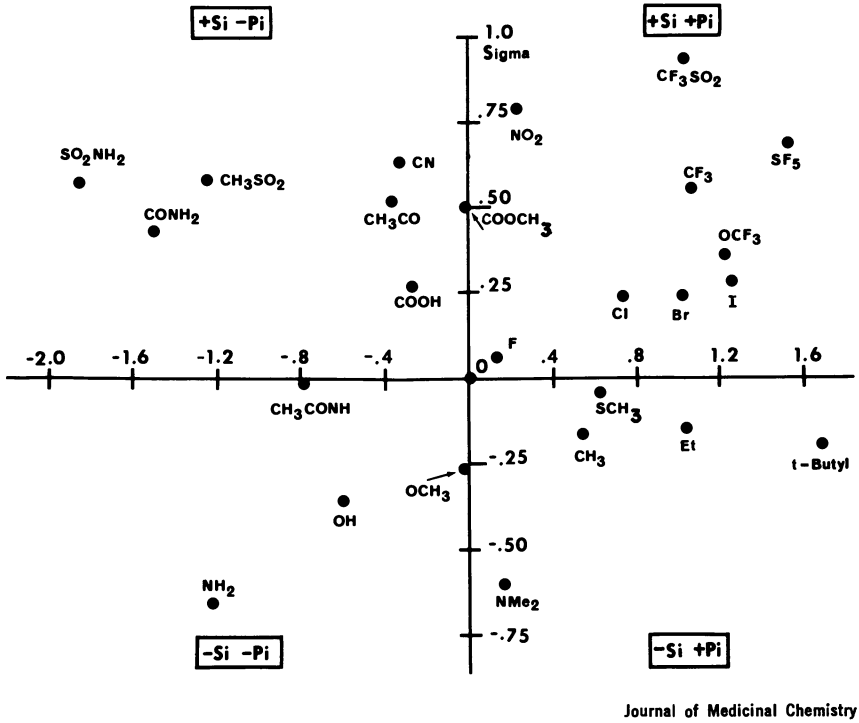


Figure 1. Two-dimensional plot of pi vs. sigma constants for aromatic substituents. The constants for the particular groups listed in Table I lie essentially on a straight line.

The substituent constants at R_3 and R_6 correlate quite well with both the pi and sigma (para) constants as shown by correlation coefficients of from 0.89 to 0.99. In hindsight, one would expect that the Hansch method should give good results using either pi or sigma constants as parameters. Actually, Hansch and I had previously obtained such correlations from the same set of data, and these results represent agreement between the two different methods. However, we were unable to decide between partition or polar factors as either one alone gave good correlations. The problem was found to reside in the particular choice of substituent groups studied; a two-dimensional plot of pi *vs.* sigma constants for aromatic substituents shows that the particular groups studied (*see* above list) lie essentially on a straight line. Therefore they cannot lead to a choice between them. This two-dimensional plot is reproduced in Figure 1; a more complete discussion of this "covariance" problem has been published (5).

When which of these methods to apply to a given set of data is considered, the choice will usually depend upon the number and type

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

of analogs which have been prepared. One should have data for at least five more compounds than the minimum required for solution of the Free-Wilson matrix. In addition, one should have two or more examples for each group at every position, if possible, to increase the confidence with which one can apply the results.

To apply the Free-Wilson method, one must have a series of closely related structures whereas the Hansch method may be applied to series of compounds with quite different structure, provided one has data for one or more physical parameters for all of the compounds in question. When one has only 8–12 compounds, only the Hansch method may be used.

Free intended this approach to be used as a guide to proper experimental design for the chemist, in planning his choice of analogs for preparation. By advance planning one can avoid the situation described above where one has only one example of a substituent at a particular position.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

Detailed Discussion of the Free-Wilson Additivity Model

In 1964, Free and Wilson introduced the method for structure–activity correlation which is based upon the assumption that each substituent group in a molecule at a specific position makes a constant additive contribution towards the overall biological activity of the molecule. One sets up a series of equations, one per compound, which mathematically expresses this concept. Solution of these multiple equations in multiple unknowns is obtained by the method of least squares, using computer techniques, and the resulting statistical parameters allow one to judge whether or not the original assumption of additivity was valid for the particular set of biological data studied. If additivity is confirmed, the model may be used to predict approximate bioactivities for those combinations of substituent groups which have not been prepared.

The ranges of substituent values at each position help identify those positions in the molecule which are most sensitive to change in substituent; these are the positions where favorable groups may be expected to increase activity appreciably. In addition, the relative activities of the substituent groups at a particular position often suggest relationships (such as with pi and sigma above) which can lead to extrapolations to suggest groups not originally studied which may be worth preparing.

Additivity—The Basic Premise. The concept of additivity of substituent group contributions is merely an expression of the medicinal chemist's intuition which has so successfully led to the development of useful therapeutic agents in the past 70 years. However, it is such a basic

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

concept that the chemist must question it and must consider its implications.

There are certainly cases where synergistic effects are seen (*3*), and these are not compatible with the concept of additivity. Many successful studies have confirmed the general concept, and the method has a built-in check in that the statistical parameters (F value, R^2, standard deviation) allow one to check the validity of the original assumption with regard to the actual set of biological data under consideration (*6*). Therefore, one is not without guidelines in using this technique, and the process can be helpful even if additivity is not confirmed, as that in itself can provide useful information. A major check of the analysis is provided when the activities of each of the compounds are calculated using the derived substituent constants. The deviations between the observed and calculated bioactivities can point out particular compounds which are poorly calculated. In an as yet unpublished case the large deviation for one compound was used to question the structure, and indeed, a rearrangement had occurred during its preparation which resulted in an incorrect assignment of its structure. The more usual cause of large deviations is in the biological test results as these are often difficult to quantitate reproducibly. Of course, a large deviation may also point towards other causes of non-additivity such as chelation effects, hydrogen bonding, or special steric considerations.

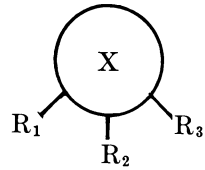
The general problem of lack of accurate reproducibility of the biological test data usually results in an unavoidable standard deviation of about 0.20 to 0.25 log $1/C$ units. Thus "additivity" really means that slight variations of this magnitude from strict additivity could not be detected.

It should be pointed out here that the Hansch method, too, assumes that each substituent plays a constant and additive role from compound to compound, and it, too, is limited by the almost irreducible standard deviation of about 0.2 or 0.25 log $1/C$ units (*8*). In a recently published paper, Cammarata treats the relationships and assumptions involved in the Hansch and Free-Wilson methods from a systematic point of view (*8*).

Procedure. The Free-Wilson method is most useful when three or more positions of a molecule are subjected to variation; although one can apply the method to cases where only two positions are involved, simple intuition can do about as well in such simple cases. As the complexity of the structural changes increases, this method becomes more and more valuable, and in very complicated systems with substituents at many positions, it can be extremely helpful. The following treatment will illustrate the method in general terms. The following generic formula repre-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

sents a family of compounds, all having a common structure X which is substituted at R_1, R_2, and R_3:



Let us assume that there are 15 compounds in this series; at R_1 we have four different groups, at R_2, five groups, and at R_3, five groups. Comparative biological data are available for all fifteen compounds. The biological data may be expressed in quantal units (*e.g.*, 0,1,2,3,4), or in activities relating them to a standard agent, or most conveniently, in terms of log $1/C$ values where C is the molar concentration of test compound which causes a standard effect. In the case of whole animal studies, C is usually expressed as moles/kg test animal. By use of log $1/C$ positive increases in value mean increased biological activity. A structural matrix is assembled in Table II.

Table II. Structural Matrix

Compound		R_1				R_2					R_3				
Number	*Log 1/C*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*
1	1.02	1	0	0	0	1	0	0	0	0	1	0	0	0	0
2	3.01	1	0	0	0	0	1	0	0	0	1	0	0	0	0
3	2.53	1	0	0	0	0	0	1	0	0	0	1	0	0	0
4	3.21	0	1	0	0	0	0	0	1	0	0	1	0	0	0
5	3.10	0	1	0	0	0	0	0	1	0	0	0	1	0	0
6	2.89	0	1	0	0	0	0	0	0	1	0	0	1	0	0
7	1.98	0	1	0	0	1	0	0	0	0	0	0	0	1	0
8	2.45	1	0	0	0	0	1	0	0	0	0	0	0	0	1
9	3.05	0	0	1	0	1	0	0	0	0	0	0	0	0	1
10	2.70	0	0	1	0	0	1	0	0	0	1	0	0	0	0
11	2.40	0	0	1	0	0	0	1	0	0	0	0	0	1	0
12	2.78	0	0	0	1	0	0	1	0	0	0	1	0	0	0
13	2.98	0	0	0	1	0	0	0	0	1	0	0	0	0	1
14	1.80	0	0	0	1	0	0	0	0	1	0	0	0	1	0
15	3.15	0	0	0	1	0	0	0	1	0	0	0	1	0	0
Number of examples		4	4	3	4	3	3	3	3	3	3	3	3	3	3

Compound 1 has group A at position R_1, group E at position R_2, and group J at position R_3. Thus the matrix defines the exact structure of each compound. The only variable numbers are the log $1/C$ values, which are experimentally determined and whose variability is the reason why so many compounds are required in excess of the theoretical number

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

required to achieve a solution. No assumptions are made about the physical parameters which may be involved, and only the exact structural data and the biological test data are used to carry out this analysis. This matrix serves as input to a computer for solution of the following equation by least squares:

$$\text{Log } 1/C = \mu + A + B + C + D + E + F + G + H + I + J + K + L + M + N$$

In this equation, μ is the over-all average log $1/C$ value for the 15 compounds, and A through N are numerical coefficients, positive and negative, which represent the contributions of each group (A through N) to the biological activity; these constants (*de novo* substituent constants, valid only for this set of data) represent the group contributions to the over-all biological activity.

The basic assumption of additivity demands that the following relationships hold for this set of data:

$$4A + 4B + 3C + 4D = 0; \text{ or } A = -B - 3/4C - D \text{ (at } R_1)$$
$$3E + 3F + 3G + 3H + 3I = 0; \text{ or } E = -F - G - H - I \text{ (at } R_2)$$
$$3J + 3K + 3L + 3M + 3N = 0; \text{ or } J = -K - L - M - N \text{ (for } R_3)$$

These are called the restrictive equations; because of these relationships between the variables there are really only three unknown terms at R_1 and four each at R_2 and R_3. Hence there are $3 + 4 + 4 = 11$ unknown terms to be solved, plus one more term for the over-all average, μ. There must be a minimum of 12 compounds to permit a solution. However, since the biological test results are the dependent variables, and each of the biological test results has a degree of uncertainty, or variability, to its value, a number of additional compounds is required to give a good degree of assurance for the results, *i.e.*, to increase the statistical significance of the derived values for each term (these are the so-called *de novo* substituent constants). It is desirable to have at least five and, preferably, 10 or more compounds in excess of the minimum required for solution. Our hypothetical sample case has only three more compounds than the 12 required and so could not be expected to give significant results.

The use of a special multiple regression analysis computer program which can include the restrictive equations properly leads directly to the proper solution. To use a standard multiple regression analysis program one must incorporate the restrictive conditions as follows. Recalling that $A = -B - 3/4C - D$, the entire column for the A term is removed from

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

the matrix; similarly, the columns for E and J are removed. Now we are left with a matrix which contains 3, 4, and 4 variables at R_1, R_2, and R_3, respectively. When introducing the expression for the compounds which contained the terms A, E, or J, one introduces the equivalent values in the same way as illustrated in Table III.

Table III. Contracted Matrix

Compound Number	*Log 1/C*	*B*	*C*	*D*	*F*	*G*	*H*	*I*	*K*	*L*	*M*	*N*
1	1.02	−1	−0.75	−1	−1	−1	−1	−1	−1	−1	−1	−1
2	3.01	−1	−0.75	−1	0	1	0	0	−1	−1	−1	−1
3	2.53	−1	−0.75	−1	0	0	1	0	0	1	0	0
4	3.21	0	1	0	0	0	1	0	0	1	0	0

Solution of the original matrix by a standard program would give incorrect statistical parameters if the restrictive equations were entered directly into the matrix. Solution of the contracted matrix in Table III by a standard program gives exactly the same results as are obtained from the original matrix by the special program of Free and co-workers. However, the substituent constants for A, E, and J must be calculated by use of the restrictive equations when the contracted matrix is used.

The least squares solution gives the values of μ and the substituent constants A through N. In addition, the F value for the over-all regression is calculated, as is the correlation coefficient, R. The term R^2 (the variance) is expressed by the (model sum of squares)/(total sum of squares), and when this value is 80% or somewhat greater, the original assumption of additivity of substituent group effects is considered to be supported. The remaining variability of 10–20% is the residual variation due to biological test variability and is almost inescapable. The F value offers an additional check of the significance of the results; this value can be compared with the decision statistic "F" value from tables to test the validity of using the derived substituent constants to calculate the biological activities of the compounds used in the Free-Wilson analysis. If the F test fails, it means that one cannot explain the different biological activities of the compounds by use of the additivity equation; in such a case one is as well off using μ as the calculated activity—a sad state of affairs indeed.

In our hypothetical example, the F value is the one for $F_{3,11}$. These subscripts derive from the analysis of variance for the regression, where $3 + 4 + 4 = 11$ degrees of freedom are attributable to the regression, and $15 - 1 = 14$ represents the total degrees of freedom in the model. Then $14 - 11 = 3$ degrees of freedom which are attributable to the error term. If the observed F value exceed the tabular value of 8.76 ($F_{3,11}$ at

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

the 5% level) or 26.13 ($F_{3,11}$ at the 1% level), the derived solution is significant at or above those levels.

In preparing the matrix for a Free-Wilson analysis, one must be careful to avoid situations which lead to singularities and hence cannot give a unique solution. The problem to be avoided is illustrated in Table IV in its simplest form.

Table IV. Situations Leading to Singularities

Compound Number	*Log 1/C*	R_1			R_2		
		A	*B*	*C*	*D*	*E*	*F*
1	1.2	1	0	0	1	0	0
2	1.4	0	1	0	0	1	0
3	1.6	0	0	1	0	1	0
4	0.9	0	1	0	0	0	1
5	0.6	0	0	1	0	1	0
6	2.1	0	1	0	0	0	1
	Total	1	3	2	1	3	2

Compound 1 in the matrix represents the unique occurrence of two substituents in one compound. This is readily seen as a violation of the medicinal chemist's cardinal rule of making only one change at a time. It is no more possible for the computer to assign values to *A* and *D* than it would be for a chemist who has changed two groups at once to ascribe the resulting change in biological activity to either one of the groups when he has no other examples of the effects of either group.

To detect such problems in advance of the computer run one should always prepare the full structure matrix, and each group which appears but once should be checked to be sure that the particular compound bearing that substituent has no other substituent which occurs only in that compound. More complex singularities, which have the same mathematical basis, are shown in Tables V and VI; these are encountered less often, but should be checked for by careful study of the original matrix.

In the latter case, although there are three examples of *A* and *D*, they occur in exactly the same set of compounds, and this is equivalent to the other cases which lead to singularities.

This explains why Hudson, Bass, and Purcell (7) obtained two different solutions to a matrix by making different substitutions using the restrictive equations. Their presentation of the matrix used is not in its most expanded form but is contracted by application of the restrictive equations; this makes it difficult to visualize the singularity problems. When their matrix is rewritten in the complete form, it becomes readily apparent that their Compounds 3 and 4 are involved in a singu-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table V. Complex Singularities

Compound Number	*Log 1/C*	*R_1*			*R_2*		
		A	*B*	*C*	*D*	*E*	*F*
1[a]	1.2	1	0	0	1	0	0
2[a]	0.7	1	0	0	1	0	0
3	0.9	0	1	0	0	0	1
4	0.8	0	1	0	0	1	0
5	1.1	0	0	1	0	0	1
6	0.3	0	0	1	0	1	0
	Total	2	2	2	2	2	2

[a] Cases leading to singularities.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

larity of the type illustrated in Table IV; their Compounds 5 and 6 have singularity problems of the type illustrated in Table I. Removal of these four compounds from the matrix leads to a more simple matrix with no such problems, and identical substituent constants result regardless of how the restrictive equations are used—*e.g.*, a unique solution is obtained.

Use of the special program developed by Free, which will accept the entire matrix, avoids this problem as the existence of a singularity prevents any solution from being obtained. Unfortunately, use of a standard regression program with application of the restrictive conditions, as already discussed, can force a solution which is one of a family of possible solutions; these are not unique, and cannot be relied upon. The technique of studying the original matrix in its entirety, to avoid singularities, will pinpoint these problems. Elimination of one or more compounds from the matrix will correct the problem.

If hydrogen is considered as one of the substituents at R_1, the resulting value of μ (the average biological activity) must be considered to be the biological activity of a hypothetical compound where R_1 is a nonentity. In this case, the substituent constant for hydrogen must be added for R_1 to obtain the value of the "unsubstituted molecule" where $R_1 = H$.

Table VI. Complex Singularities

Compound Number	*Log 1/C*	*R_1*			*R_2*		
		A	*B*	*C*	*D*	*E*	*F*
1[a]	0.9	1	0	0	1	0	0
2[a]	0.7	1	0	0	1	0	0
3	1.3	0	1	0	0	1	0
4[a]	1.2	1	0	0	1	0	0
5	1.1	0	1	0	0	0	1
6	0.3	0	0	1	0	1	0
	Total	3	2	1	3	2	1

[a] Cases leading to singularities.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

To avoid this, and to place the resulting substituent constants on a scale relative to hydrogen equal to zero, Cammarata (*4*) used the technique of setting the substituent constant for hydrogen as zero by not including hydrogen as one of the groups in the matrix—*e.g.*, in Table V, neither A,B,C, nor D,E, or F is hydrogen. In this case, the compound where R_1 and R_2 are hydrogen would be entered as follows: log $1/C$ = 0 0 0 0 0 0.

An additional advantage of this approach is that now one need not add the restrictive equation since one variable (H) has been removed from the matrix for each position. Thus a standard multiple regression program can now be directly employed, taking care to avoid the singularity problem, of course. Should hydrogen not occur as a substituent at one of the positions, one may arbitrarily set another substituent group at zero as was illustrated above for hydrogen.

Predictive Use of Free-Wilson Substituent Constants. The newly found substituent constants are lined up in decreasing order of activity for each position. A prediction may be made for all possible compounds arising from chemically allowable combinations of one group at each position. In these predictions it must be remembered that constants which were obtained for groups which occurred only once or twice in the matrix usually have a low degree of significance. Here one should be guided by reference to a *T* test value to establish the significance of the difference between the substituent constants at each position (*6*). Of course, it is possible to build the calculations for all possible combinations of the substituents at each position into a computer program, and this results in a listing of predicted biological activities for all possible compounds.

Such predictions must be used with caution but can be good guides when the statistical parameters are favorable. One must avoid the prediction of extreme activity for unusually highly substituted analogs, which might be almost impossible to prepare.

For the model illustrated in Table II, $4 \times 5 \times 5$ or 100 analogs are exemplified; from a study of from 20 to 25 compounds we should have a good idea of the activities to be expected for the other 75 to 80 analogs. The higher the standard deviation for the analysis, the lower the accuracy of the predicted values will be. For typical biological tests, variation of from one-half to twice the observed activity is about average.

Extrapolations beyond the confines of the exact substituent groups studied may be made if, after a study of the list of substituent group values at each position, correlations with physical constants are noticed. Such correlations may be observed by graphical procedures, or by the use of regression analysis. The use of new groups not yet studied can now be suggested by reference to tables of physical constants. Extra-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch008

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

polation beyond the limits of the range of values for those groups already studied should be carefully considered, and the resulting predictions should not be expected to be very accurate. However, by this method a Free-Wilson analysis can lead to suggestions for new compounds in a manner similar to use of the Hansch method.

Literature Cited

1. Hansch, C., *Accounts Chem. Res.* (1969) **2,** 232.
2. Free, S. M., Jr., Wilson, J. W., *J. Med. Chem.* (1964) **7,** 395.
3. Fujita, T., Ban, T., *J. Med. Chem.* (1971) **14,** 148.
4. Cammarata, A., Yau, S. J., *J. Med. Chem.* (1970) **13,** 93.
5. Craig, P. N., *J. Med. Chem.* (1971) **14,** 680.
6. Snedecor, G. W., "Statistical Methods," Iowa State University Press, Ames, Iowa, 1966.
7. Hudson, D. R., Bass, G. E., Purcell, W. P., *J. Med. Chem.* (1970) **13,** 1184.
8. Cammarata, A., *J. Med. Chem.* (1972) **15,** 573.

RECEIVED June 17, 1971. Work done at Smith Kline and French Laboratories, Philadelphia, Pa.

9

Structure–Activity Relations

II. Antibacterial Activity of 3-Benzoylacrylic Acids and Esters

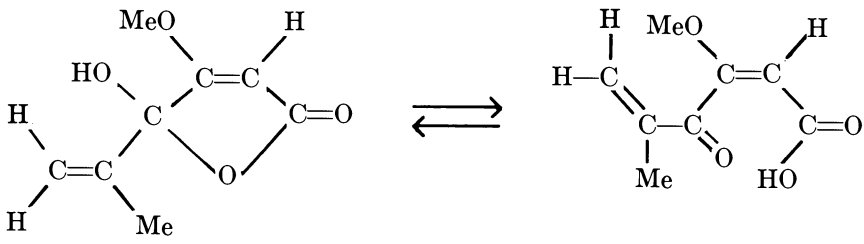
KEITH BOWDEN and M. P. HENRY

The University of Essex, Colchester, Essex, England

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

The bacteriostatic activity of a series of substituted trans-*3-benzoylacrylic acids has been successfully correlated by Hansch linear free energy relations involving polar and partition substituent constants. The activity–lipophilicity relations for this series closely parallel those found previously for other antibacterial agents, with an ideal lipophilic character for gram-positive cells of 6.1 and, for linear dependence, a slope of 0.7. A polar reaction constant, ρ, of about −0.6 to −0.7 is given. A possible mode of action for these acids and their related substituted* cis- *and* trans-*3-benzoyl acrylic acids and esters is discussed as an enzyme-inhibitor interaction.*

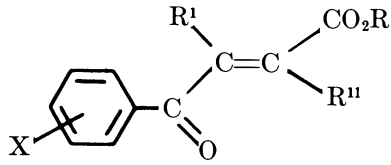
The presence of the ketovinyl groups has been implicated in the antibacterial activity of a number of substances (*1*), both natural and synthetic (*1–7*). An example is penicillic acid which exists mainly in the cyclic form (I) whereas, at higher pH, the anion of the chain form (II) occurs (*6, 8*). Addition to or loss of the reactive double-bond in these compounds markedly reduces their activity (*1, 6, 7*). The bacteriostatic



I II

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

and fungicidal activities of many synthetic compounds containing the ketovinyl grouping have been investigated (*1–7*). A systematic study of the bacteriostatic activity of a series of substituted *trans*-3-benzoylacrylic acids (III, R = R^1 = R^{11} = H) has been made by Kirchner, Bailey, and Cavallito (*7*).



III

Hansch *et al.* (*9, 10*) have developed linear free energy relations which can correlate biological activity and chemical structure. Equations 1 and 2, together with simplified versions, have been proposed to relate the molar concentration, C_x, of a substituted compound of a series, which all cause an equivalent biological response, to the hydrophobic bonding or partition constant, π, and the Hammett constant, σ. The constants a, b,

$$\log (1/C_x) = -a\pi^2 + b\pi + \rho\sigma + c \quad (1)$$

$$\log (1/C_x) = b\pi + \rho\sigma + c \quad (2)$$

ρ, and c are obtained from the regression analysis and define the response of the biological system to structural features. Hansch *et al.* (*11*) have studied the bacteriostatic activity of diverse compounds. These activities were normally well correlated by Equation 1 or 2, usually without a polar contribution. The ideal partition character for optimum activity, log P_o, was found to be about 4 for gram-negative and about 6 for gram-positive bacteria (Equation 1); however, where a linear relation (Equation 2) resulted, a slope of b equal to about 0.7 was observed. These results were related to the cell wall composition of the bacteria. Hansch has made two studies (*12, 13*) of the structure–activity relations in penicillin derivatives. An earlier attempt had been made using the simple Hammett relation (*14*). The first of Hansch's studies (*12*) gave good correlations for the *in vivo* testing of a series of substituted penicillins having the side chain $PhOCH(CH_2)_nCH_3$ while those correlations for *in vitro* activity were less satisfactory. The lipophilic character of the substituent appeared to be more important and gave a linear relation between partition factors and activity (Equation 2) having b equal to about -0.4_5. The second study (*13*) reported a correlation of the activity of substituted 2,6-dialkoxyphenyl penicillins against a re-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table I. Bacteriostatic Activities and Substituent Constants for Substituted *trans*-3-Benzoylacrylic Acids[a]

	Substituent	*log* (1/C)[b] *v.* Staphylococcus aureus 209	*v.* Cl. perfringens	*v.* M. tuberculosis H37Rv	σ	π
1	H	0.12_5	-0.09_7	-0.39_8	0.0	0.0
2	4′-Me	0.30_1	-0.09_7	0.12_5	−0.17	0.43
3	4′-CH_2Me	1.00_0	0.42_6	0.30_1	−0.15	0.86
4	4′-$(CH_2)_2Me$	1.30_1	0.42_6	1.00_0	−0.15	1.29
5	4′-$(CH_2)_3Me$	2.00_0	0.90_3	1.30_1	−0.15	1.72
6	4′-$(CH_2)_4Me$	2.12_5	0.90_3	2.00_0	−0.15	2.15
7	4′-$(CH_2)_5Me$	2.30_1	0.42_6	2.30_1	−0.15	2.58
8	4′-$(CH_2)_6Me$	2.60_2	2.30_1	3.00_0	−0.15	3.01
9	4′-$(CH_2)_7Me$	3.12_5	2.42_6	3.30_1	−0.15	3.44
10	4′-$(CH_2)_8Me$	2.60_2	2.60_2	3.00_0	−0.15	3.87
11	4′-$(CH_2)_9Me$	2.60_2	3.60_2	3.12_5	−0.15	4.30
12	4′-$CHMe_2$	1.30_1	1.60_2	0.30_1	−0.15	1.40
13	4′-CMe_3	1.60_2	1.90_3	0.30_1	−0.20	1.78
14	4′-OMe	0.00_0	0.60_2	0.00_0	−0.27	0.11
15	4′-OCH_2Me	0.60_2	0.60_2	0.00_0	−0.27	0.54
16	4′-$O(CH_2)_5Me$	1.60_2	1.30_1	1.30_1	−0.27	2.26
17	4′-Cl	0.60_2	0.42_6	0.00_0	+0.23	0.80
18	4′-Br	0.60_2	0.42_6	0.12_5	+0.23	1.01
19	3′-NO_2	-0.69_9	-0.39_8	-0.69_9	+0.71	0.10

[a] Substituent constants are either literature values (*18*, *19*) or values estimated from the latter.
[b] The *C* values are literature (*7*) values in M^{-1}.

sistant *S. aureus* strain. Using Equation 2, *b* was about -0.2_5, and polar factors were now significant, with ρ equal to about 1.8.

In this study we report a Hansch analysis of the activity of a series of substituted *trans*-3-benzoylacrylic acids. The bacteriostatic and fungistatic activity of several series of substituted *cis*- and *trans*-3-benzoylacrylic acids and their methyl esters are given. The structure–activity relations observed are discussed.

Experimental

Acids and Methyl Esters. These compounds have been prepared, and their syntheses are described elsewhere (*15*). After careful purification, their physical constants agreed well with literature values, or they had satisfactory elemental analyses.

Partition Constants. The partition coefficient of *trans*-3-benzoylacrylic acid between water and 1-octanol was measured in the manner described previously (*16*) as *P* equal to 34 ± 3.

Antibacterial and Antifungal Activities. The activities of the compounds listed in Table V were measured against nine bacteria and five

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

fungi. The results shown for *S. aureus NCTC 7447, M. tuberculosis Inh. H37 Rv,* and *Microsporum canis ATCC 10214* were considered to be typical. The activities were determined by dissolving the compounds in a suitable solvent at different concentrations. The assay procedure then used was standard (*17*), and the minimum inhibitory concentrations were determined as the lowest concentration at which no growth occurred. Controls containing the same amount of solvent were included in each experiment. These measurements were carried out in the laboratories of J. R. Geigy, A. G. Basel, Switzerland.

Results

The bacteriostatic activities of a series of substituted *trans*-benzoylacrylic acids against *S. aureus 209, Cl. perfringens,* and *M. tuberculosis H37Rv* are shown in Table I, together with the relevant π and σ values. The activities are from the study of Kirchner *et al.* (*7*) while the substituent constants are either literature values (*18, 19*) or values estimated from the latter. Regression analyses were made using multiple regression analysis (least-squares) using Equation 1 and 2 and simplified versions. In addition to the analyses using all the substituents, a limited series (substituents 1–3, 12–19) was also studied to eliminate the possible overweighting from the polar contributions of the long chain alkyl substituents (substituents 4–11). The analyses are shown in Tables II–IV. The bacteriostatic and fungistatic activities for several series of substituted *cis*- and *trans*-3-benzoylacrylic acids and their methyl esters for three typical species, *S. aureus NCTC8, M. tuberculosis Inh. H37Rv,* and *Microsporum canis ATCC1* are shown in Table V.

Discussion

The correlations shown in Tables II–IV show successful results. The best correlation for activity against *S. aureus 209* using the complete series (Table II, regression 3) is shown in Equation 3.

Table II. Correlations of Bacteriostatic Activity of the Substituted *trans*-3-Benzoylacrylic Acids[a]

v. *Staphylococcus aureus 209*

					r	s	n
1	$-0.166\pi^2(3.85)$	$+ 1.424\pi(7.88)$		$- 0.294$	0.966	0.294	19
2		$+ 0.700\pi(10.70)$	$- 0.952\sigma(2.63)$	$+ 0.110$	0.953	0.340	19
3	$-0.142\pi^2(3.67)$	$+ 1.285\pi(7.70)$	$- 0.696\sigma(2.48)$	$- 0.222$	0.976	0.255	19
4[b]		$+ 0.770\pi(7.55)$	$- 0.765\sigma(3.15)$	$- 0.032$	0.961	0.219	11
5[b]	$-0.262\pi^2(2.36)$	$+ 1.325\pi(5.32)$	$- 0.784\sigma(4.05)$	$- 0.188$	0.979	0.174	11

[a] n = number of compounds; r = correlation coefficient; s = standard deviation, and the quantity in parentheses is the student's t test (*20*) for the significance of the regression variable.

[b] Limited series with substituents 1–3, 12–19 (Table I).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table III. Correlations of Bacteriostatic Activity of the Substituted *trans*-3-Benzoylacrylic Acids[a]

v. *Cl. perfringens*

					r	s	n
6		$+ 0.724\pi(7.79)$		$- 0.138$	0.884	0.515	19
7	$+0.092\pi^2(1.24)$	$+ 0.355\pi(1.14)$		$+ 0.073$	0.895	0.507	19
8		$+ 0.691\pi(6.97)$	$- 0.538\sigma(0.98)$	$- 0.126$	0.891	0.516	19
9	$+0.117\pi^2(1.58)$	$+ 0.206\pi(0.64)$	$- 0.757\sigma(1.39)$	$+ 0.149$	0.907	0.493	19
10[b]		$+ 0.717\pi(3.92)$	$- 0.710\sigma(1.63)$	$- 0.018$	0.874	0.393	11

[a, b] For footnotes *see* Table II.

$$\log (1/C) = -0.142\pi^2 + 1.285\pi - 0.696\sigma - 0.222 \quad (3)$$

The correlation is highly statistically significant, and the contribution of the polar term is confirmed by its increased significance for the limited series (Table II, regression 5). The calculated and observed values of log $(1/C)$ agree to a mean value of ± 0.174 (Equation 3). The optimum partition constant (*21*) of π_o from Equation 3 equals 4.5_2. A value of log P for *trans*-3-benzoylacrylic acid has been found to be 1.5_4 (*see* Experimental), and the ideal lipophilic character (log P_o) for this series therefore equals 6.0_6. This is in remarkable agreement with those log P_o values found by Hansch *et al.* (*11*) for several different antibacterial agents against S. *aureus* (in the range 5.2 to 6.4).

The most successful correlation for activity against *Cl. perfringens* using the complete series (Table III, regression 6) is shown in Equation 4.

$$\log (1/C) = 0.724\pi - 0.138 \quad (4)$$

The other relations do not demonstrate any highly significant contributions from other factors. The limited series does indicate greater importance for polar effects than that observed for the complete series but does not reach high significance. The calculated and observed values agree to a mean value of ± 0.367 (Equation 4). Although this is less successful than the previous correlation for S. *aureus*, the slope, *b*, of about 0.7 is quite close to that of the same type of relation for S. *aureus* (Table II, regression 2 or 7). Further, both correlations give slopes close to the mean value of 0.73 found by Hansch (*11*) (range 0.52 to 0.91) for a number of series of antibacterials against gram-positive bacteria. The most successful correlation for activity against *M. tuberculosis* using the complete series (Table IV, regression 11) is shown in Equation 5,

$$\log (1/C) = 0.962\pi - 0.530 \quad (5)$$

with calculated and observed values agreeing to a mean value of ± 0.341. However, in this case, the limited series indicates a statistically significant

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

contribution from polar factors (Table IV, regression 15), shown in Equation 6,

$$\log (1/C) = 0.500\pi - 0.589\sigma - 0.316 \quad (6)$$

with calculated and observed values agreeing to a mean value of ±0.141. Although there appears, in general, to be a rather limited, direct relation between π and σ (*19*), this is not the cause of the polar contribution in this case. It therefore seems, from Equations 3 and 6, that the antibacterial activity of the series under study does have a polar substituent effect dependence. The reaction constant, ρ, of about −0.6 to −0.7 indicates increased antibacterial activity with increased electron-releasing capacity of the substituent.

The bacteriostatic and fungistatic activities of several series of substituted *cis-* and *trans*-3-benzoylacrylic acids and their methyl esters are shown in Table V. These results are not as diverse in partition factors as those of Kirchner *et al.* (*7*). Although unsuitable for Hansch analysis, our results show the same trends as those analyzed, and it is possible to state a number of general conclusions. Substitution of a methyl group at the 2 or 3 position markedly reduces the bacteriostatic or fungistatic activity. In general, the methyl esters are more active than the acids. Both of these effects are not those that would be estimated for the changes in lipophilicity for the introduction of these methyl groups (assuming additivity of partition factors, with $\Delta\pi$ equal to +0.3 to +0.5 (*19*, *22* *23*) and the applicability of Equations 3 or 5. The methyl esters are about 10 times more active and the 2- or 3-methyl acids are about 1/20th as active as would be expected. However, when the activities are measured at pH 7–8, the acids will be present as the carboxylate anions. This will considerably alter their partition characteristics, in comparison with the methyl esters, and make any direct interpretation impossible. As bacteriostats, there is little difference between trans- and cis-isomers whereas fungistats the cis-isomers are more active.

Table IV. Correlations of Bacteriostatic Activity of the Substituted *trans*-3-Benzoylacrylic Acids[a]

v. *M. tuberculosis Inh. H37 Rv*

					r	s	n
11		$+ 0.962\pi(12.60)$		$- 0.530$	0.950	0.423	19
12	$+0.0001\pi^2(0.00)$	$+ 0.965\pi(3.69)$		$- 0.555$	0.952	0.429	19
13		$+ 0.934\pi(11.49)$	$- 0.444\sigma(0.99)$	$- 0.520$	0.953	0.423	19
14	$+0.007\pi^2(0.10)$	$+ 0.907\pi(3.18)$	$- 0.455\sigma(0.95)$	$- 0.504$	0.953	0.437	19
15[b]		$+ 0.500\pi(5.10)$	$- 0.589\sigma(2.52)$	$- 0.316$	0.926	0.210	11
16[b]	$+0.106\pi^2(0.77)$	$+ 0.277\pi(0.90)$	$- 0.581\sigma(2.42)$	$- 0.254$	0.932	0.216	11

[a, b] For footnotes *see* Table II.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

Table V. Bacteriostatic and Fungistatic Activities for Substituted *cis*- and *trans*-3-Benzoylacrylic Acids and Their Methyl Esters

Series	*Minimum Inhibitory Concentrations*[a] ($10^3 M$)		
	S. aureus NCTC 7447	M. tuberculosis Inh. H37 Rv	Microsporum canis ATCC 10214
3-Benzoylacrylic acid			
trans-H	5.5	0.5_5	1.7
trans-4′-Me	1.6	0.5_5	$<0.1_6$
trans-4′-OMe	5.0	0.5_0	1.5
trans-4′-F	5.0	1.0	1.5
trans-4′-Cl	1.4	0.5_0	0.3_0
trans-4′-Br	1.2	0.8_0	0.4_0
trans-4′-I	$>0.3_0$	$>0.3_0$	0.3_0
trans-3′-NO_2	15	1.5	5.0
Methyl 3-benzoylacrylate			
trans-H	0.1_6	0.05_0	0.05_0
trans-4′-Me	0.1_6	0.05_0	0.05_0
trans-4′-OMe	0.1_4	0.05_0	$<0.005_0$
trans-4′-F	0.1_6	0.05_0	0.005_0
trans-4′-Cl	0.1_3	0.04_0	0.04_0
trans-4′-Br	0.1_1	0.4_0	$<0.004_0$
trans-4′-I	0.09_0	0.03_0	0.03_0
trans-3′-NO_2	0.4_0	0.4_0	0.04_0
cis-H	0.1_6	0.05_0	$<0.005_0$
cis-4′-Me	0.5_0	0.05_0	$<0.005_0$
cis-4′-MeO	0.1_4	0.05_0	$<0.005_0$
cis-4′-F	0.1_4	0.05_0	$<0.005_0$
cis-4′-Br	0.4_0	0.04_0	$<0.004_0$
cis-4′-I	0.09_0	0.03_0	$<0.003_0$
3-Benzoyl-2-methylacrylic acid			
trans-H	18	$>0.5_5$	$>0.5_5$
trans-4′-Me	16	$>0.5_0$	$>0.5_0$
trans-4′-OMe	15	$>0.4_5$	$>0.4_5$
trans-4′-Cl	13	$>0.4_5$	$>0.4_5$
trans-4′-Br	12	$>0.3_5$	$>0.3_5$
cis-4′-Me	15	$>0.5_0$	$>0.5_0$
cis-4′-OMe	15	$>0.4_5$	$>0.4_5$
cis-4′-Cl	15	$>0.4_5$	$>0.4_5$
cis-4′-Br	$>0.3_5$	$>0.3_5$	$>0.3_5$
Methyl 3-Benzoyl-2-methylacrylate			
trans-4′-Me	>15	0.4_5	0.4_5
trans-4′-OMe	1.4	0.4_5	0.4_5
trans-4′-Cl	0.4_c	$>0.4_0$	0.04_0
trans-4′-Br	>12	0.3_5	0.03_5

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

Table V. Continued

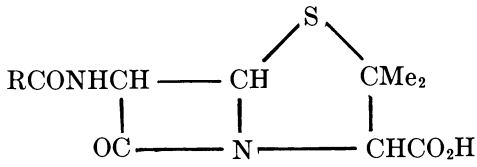
Series	*Minimum Inhibitory Concentrations*[a] (10^3M)		
	S. aureus NCTC 7447	M. tuberculosis Inh. H37 Rv	Microsporum canis ATCC 10214
3-Benzoyl-3-methylacrylic acid			
trans-H	18	$>0.5_5$	$>0.5_5$
trans-4′-Me	15	$>0.5_0$	$>0.5_0$
trans-4′-OMe	14	$>0.4_5$	$>0.4_5$
trans-4′-Cl	$4._5$	$>0.4_5$	$>0.4_5$
trans-4′-Br	12	$>0.3_5$	$>0.3_5$
trans-3′-NO_2	13	$>0.4_5$	$>0.4_5$
cis-H	>16	$5._5$	16
cis-4′-Me	>15	$5._0$	$5._0$
cis-4′-MeO	$>4._5$	$>4._5$	$>4._5$
cis-4′-Cl	>13	$4._5$	$4._5$
cis-4′-Br	>11	$1._0$	$3._5$
cis-4′-NO_2	$>0.4_5$	$>0.4_5$	$>0.4_5$
Methyl 3-Benzoyl-3-methylacrylate			
trans-H	$1._5$	0.5_0	0.05_0
trans-4′-Me	15	0.4_5	0.4_5
trans-4′-Cl	>13	0.04_0	0.04_0
trans-4′-Br	>12	0.3_5	0.03_5
trans-3′-NO_2	0.4_0	0.04_0	0.04_0
cis-H	>16	$>0.5_0$	$>0.5_0$
cis-4′-Me	>15	$>0.4_5$	$>0.4_5$
cis-4′-OMe	$>0.4_5$	$>0.4_5$	$>0.4_5$
cis-4′-Cl	>14	$>0.4_0$	$>0.4_0$
cis-4′-Br	>12	$>4._5$	$>4._5$

[a] *See* Experimental for details.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

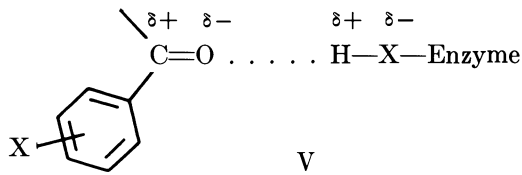
The reason for the activity and mode of action of these ketovinyl compounds as antibacterials has been discussed (*1–7*). The activity of these and related compounds has been associated with the reactivity of the activated double bond towards nucleophilic, biologically essential thiol and/or amino groups. This is in accord with studies of the reactivity of these compounds with nucleophilic reagents (*24, 25*). A 2- or 3-methyl group is known to decrease the reactivity of the double-bond towards nucleophiles (*26*). Although this is in agreement with the lower bacteriostatic activity of the methylacrylic acids, the reaction constant for substitution in the benzoyl group (Equations 3 and 6) is of the opposite sign to that expected if facilitating nucleophilic addition (*27*). An interesting structural analogy can be made between the powerful antibiotic penicillins (IV) and the bacteriostatic 3-benzoylacrylic acids (III). Three

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.



IV

essential and highly important structural characteristics of an active penicillin are the reactive β-lactam, the side-chain carboxylic acid, and the acylamino groups (*28*). Comparable sites, both in the stereochemical and reactivity sense, exist in the 3-benzoylacrylic acids. The correspondence of the "terminal" carboxylic and acyl groups is obvious. Moreover both the "central" groups are highly reactive towards nucleophiles (*27, 29*). Studies (*30, 31, 32*) of the mode of action of penicillins indicate their ability to interact with and inactivate an essential enzyme whose substrate it resembles. The active site of this enzyme has probably three points of interaction with the penicillin (*30*)—*i.e.*, those three features of penicillin already indicated. It is possible to speculate that 3-benzoylacrylic acids act in a similar but much less effective manner as an enzymatic inhibitor. This enzyme would probably have a structure closely related to that inactivated by penicillins. The reaction constant found in this study could indicate polar facilitation of an interaction such as hydrogen bonding (*33*) from the enzyme site to the aroyl carbonyl group (V). However, Hansch (*13*) found a reaction constant of the opposite



V

sign for polar substitution in a series of 2,6-dialkoxyphenyl penicillins. Unfortunately this is a correlation of their activity against a resistant bacterial strain, and it is likely that the activity of this series of penicillins depends on the inhibition of an interaction at this acylamino group by the "bulky" ortho substituents (*34*). Analyses of the activity of simple penicillins (*12, 35*) have not demonstrated dependence on the polar substituent effect although a study of a system closely comparable with the 3-benzoylacrylic acids has not yet been made.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Acknowledgments

We are grateful to W. Hoyle and H. F. Ridley of Geigy (U.K.) Ltd., Trafford Park, Manchester, for discussions and for arranging for the measurements of the antibacterial and antifungal activities, and to Geigy (U.K.) Ltd. for financial support of this work.

Literature Cited

1. Geiger, W. B., Conn, J. E., *J. Amer. Chem. Soc.* (1945) **67,** 112.
2. Papa, D., Schwenk, E., Villani, F., Klingsberg, E., *J. Amer. Chem. Soc.* (1948) **70,** 3356.
3. Rinderknecht, H., Ward, J. L., Bergel, F., Morrison, A. L., *Biochem. J.* (1947) **41,** 463.
4. Worrall, R. L., *Med. World* (1946) 2; *C.A.* (1947) **41,** 6598.
5. Haynes, L. J., *Quart. Rev.* (1948) **2,** 46.
6. Cramer, B. J., Schroeder, W., Moran, W. J., Nield, C. H., Edwards, M., Jarowski, C. I., Puetzer, B., *J. Amer. Pharm. Ass.* (1948) **37,** 439.
7. Kirchner, F. K., Bailey, J. H., Cavallito, C. J., *J. Amer. Chem. Soc.* (1949) **71,** 1210.
8. Birkenshaw, J. H., Oxford, A. E., Raistrick, H., *Biochem. J.* (1936) **30,** 394.
9. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **86** 1616.
10. Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., Streich, M., *J. Amer. Chem. Soc.* (1963) **85,** 2817.
11. Lien, E. J., Hansch, C., Anderson, S. M., *J. Med. Chem.* (1968) **11,** 430.
12. Hansch, C., Steward, A. R., *J. Med. Chem.* (1964) **7,** 691.
13. Hansch, C., Deutsch, E. W., *J. Med. Chem.* (1965) **8,** 705.
14. Hansen, O. R., *Acta Chem. Scand.* (1962) **16,** 1593.
15. Bowden, K., Henry, M. P., *J. Chem. Soc. B* (1971) 156; *J. Chem. Soc. Perkin II* (1972) **201,** 209.
16. Bowden, K., Young, R. C., *J. Med. Chem.* (1970) **13,** 225.
17. Grove, D. C., Randall, W. A., "Assay Methods of Antibiotics," Medical Encyclopedia, Inc., New York, 1955.
18. McDaniel, D. H., Brown, H. C., *J. Org. Chem.* (1958) **23,** 420.
19. Fujita, T., Iwasa, J., Hansch, C., *J. Amer. Chem. Soc.* (1964) **86,** 5175.
20. Snedecot, G. W., "Statistical Methods," 5th ed., Iowa State University Press, Ames, 1956.
21. Hansch, C., Steward, A. R., Anderson, S. M., Bentley, D., *J. Med. Chem.* (1967) **11,** 1.
22. Iwasa, J., Fujita, T., Hansch, C., *J. Med. Chem.* (1965) **8,** 150.
23. Hansch, C., Anderson, S. M., *J. Org. Chem.* (1967) **32,** 2583.
24. Nesmeyanov, A. N., Rybinskaya, M. I., Rybin, L. V., *Russian Chem. Rev.* (1967) **36,** 453.
25. Shenhav, H., Rappoport, Z., Patai, S., *J. Chem. Soc. B* (1970) 469.
26. Suminov, S. I., Kost, A. N., *Russian Chem. Rev.* (1969) **38,** 884.
27. Crowell, T. I., Helsley, G. C., Lutz, R. E., Scott, W. L., *J. Amer. Chem. Soc.* (1963) **85,** 443.
28. Doyle, F. P., Nayler, J. H. C., *Advan. Drug Res.* (1964) **1,** 1; *see* section 8.1, p. 41.
29. Clarke, H. T., Johnson, J. R., Robinson, R., Eds., "The Chemistry of Penicillin," Princeton University Press, Princeton, 1949.
30. Richmond, M. H., "Biochemical Studies of Antimicrobial Drugs," B. A. Newton, P. E. Reynolds, Eds., p. 301, Cambridge University Press, London, 1966.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

31. Park, J. T., "Biochemical Studies of Antimicrobial Drugs," p. 70, Cambridge University Press, London, 1966.
32. Tipper, D. J., Strominger, J. L., *J. Biol. Chem.* (1968) **243,** 3169 and subsequent papers.
33. Taft, R. W., Gurka, D., Joris, L., Schleyer, P. von R., Rakshyo, J. W., *J. Amer. Chem. Soc.* (1969) **91,** 4801.
34. Ref. *28,* section 8.5, p. 51.
35. Biagi, G. L., Guerra, M. C., Barbaro, A. M., Gamba, M. F., *J. Med. Chem.* (1970) **13,** 511.

RECEIVED May 10, 1971.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch009

10

Structure–Activity Correlations of Acaricidal Hydrazones—Uncouplers of Oxidative Phosphorylation

K. H. BÜCHEL and W. DRABER

Bayer AG, 56 Wuppertal 1, Postfach 13 01 05, Germany

α-Cyanocarbonylphenylhydrazones are a group of compounds which contain effective acaricides and insecticides. Their mode of action is the uncoupling of oxidative phosphorylation. pI_{50} values from rat liver mitochondria were correlated with physicochemical parameters by Hansch equations. The set consisted of 60 compounds with substituent variations in six positions of the molecule. The pI_{50} values ranged from 7.09 to 4.96. Experimental compound data (pK_a, log P, R_M) and extrathermodynamic substituent parameters (π, σ, E_s) have been used in the regression analyses. The integral parameters pK_a and log P gave correlations of only moderate significance. Therefore, it was necessary to separate the substituents. The multiparameter equations obtained in this way gave good agreement between found and calculated pI_{50} data. Electronic, hydrophobic, and, to a lesser degree, steric substituent effects contribute to the uncoupling activity. The significance of these results in relation to current theories on the mechanism of oxidative phosphorylation is discussed.

The search for insecticides with modes of action different from the well-known acetylcholinesterase inhibition led us to uncouplers of oxidative phosphorylation (*1, 2*). An inherent advantage of such pesticides would be the absence of cross-resistance with organophosphorus compounds and chlorinated hydrocarbons. The number of commercial pesticides which are likely to act by uncoupling of oxidative phosphorylation is small. All of them can be regarded as derivatives of the

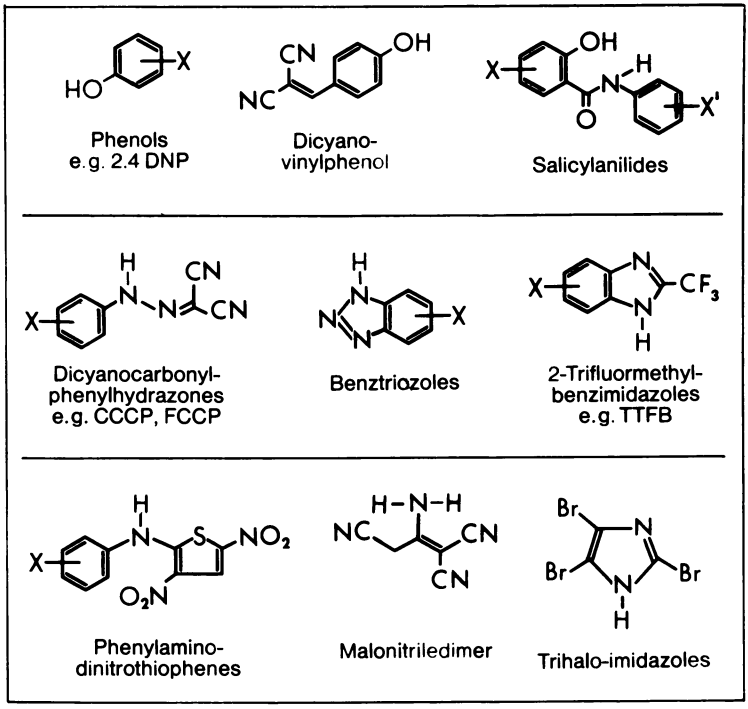


Figure 1. Uncouplers of oxidative phosphorylation

classical uncoupler 2,4-dinitrophenol (*3*), and of NH acidic benzimidazoles (*4, 5*). Several other classes of experimental compounds are known which affect ATP formation in mitochondria from different sources in the same way as dinitrophenol, some of them at low concentrations. They can be divided roughly into two groups: (1) acidic OH, and (2) acidic NH compounds (Figure 1).

From earlier work with benzimidazoles and several other types of NH acidic heterocycles we had learned (*6*) that by trying to use uncouplers as pesticides or more specifically as insecticides and acaricides, one has to cope with two problems. First, since mitochondrial phosphorylation is a ubiquitous process, both pests and mammals may be harmed; secondly, the energy producing pathways in mitochondria and chloroplasts seem to be similar; hence some of the uncouplers are also toxic to plants (*7*).

The α-Acyl-α-cyanocarbonylphenylhydrazones

To overcome these disadvantages, we looked for a system which seemed to be more liable to biochemical breakdown in mammals and showed low phytotoxicity—the α-acyl-α-cyanocarbonylphenylhydrazones

(Figure 2) (*8*). These compounds are generally much less toxic to mammals and less phytotoxic than the related dicyanocarbonylphenylhydrazones investigated by DuPont workers several years ago (*9*). Some of our hydrazones (Figure 2) show interesting insecticidal and acaricidal properties (not discussed in detail here). Many are potent uncouplers of oxidative phosphorylation in rat liver mitochondria.

We chose 60 compounds with pI_{50} values ranging from 7.1 to 4.9 and subjected them to regression analysis using several physicochemical parameters (Table I). The 60 compounds contained variations in six positions of the basic structure. Quantitative structure–activity correlations with as many individual uncouplers in one equation have not yet been published. As far as we know, the Hansch approach has been applied to uncouplers of oxidative phosphorylation only twice: first in 1965 by Hansch and co-workers to phenols and recently by Muraoka and Terada to *N*-phenylanthranilic acids. From Muraoka's data we recalculated the correlation with π and σ and obtained an equation which gave the best fit (last equation, Figure 3).

As the equations show, linear correlations with the variables π and σ gave satisfactory results. This is certainly a simplification resulting from limited variance in the substituents. One would assume that square terms of the hydrophobic parameter are necessary in every correlation with biological activity not only to account for the "random walk" penetration process as in the original derivation of his equation by Hansch, but also, or even predominantly, as a description of the fact that numerous indifferent hydrophobic sites within the biological system compete with the site of action for the active molecule. In a first attempt we calculated regression equations for our hydrazones with the molecular parameter

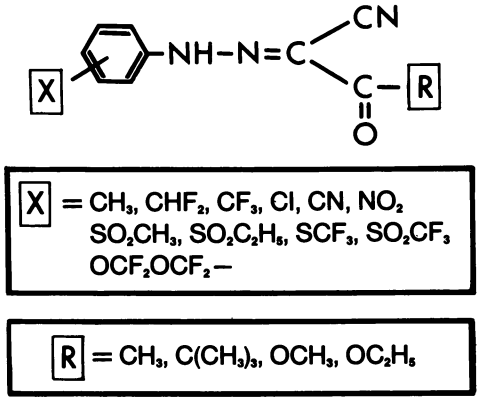
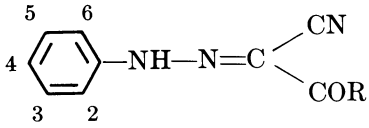


Figure 2. Structure of the α-acyl-α-cyanocarbonylphenylhydrazones

pK, which describes the electronic effects on the dissociation of the NH group, and log P or R_M which account for the lipophilicity of the molecule as a whole. The choice of pK as a parameter tacitly implies a hypothesis on the toxophore within the molecule—the assumption that the NH group is essential. This appears justified since the acidic NH, in connection with an aromatic or heterocyclic system, is the only feature common to a number of uncouplers (*see* Figure 1). Additional proof comes from the fact that the hydrazones lose their *in vitro* uncoupling activity when the NH group is alkylated.

Table I. Physicochemical Constants

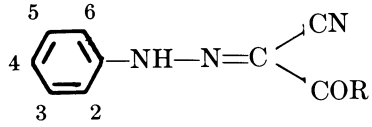


No.	X_2	X_3	X_4	X_5	X_6	R
1	H	H	OCF_2–	O-CF_2	H	*t*-Bu
2	H	CF_3	H	CF_3	H	*t*-Bu
3	H	Cl	Cl	Cl	H	OEt
4	H	Cl	H	Cl	H	*t*-Bu
5	H	CF_3	H	CF_3	H	OMe[a]
6	H	Cl	H	Cl	H	OEt
7	H	H	OCF_2–	OCF_2	H	OMe
8	H	CF_3	H	CF_3	H	OMe [b]
9	H	CF_3	Cl	H	H	OEt
10	H	Cl	Cl	Cl	H	OMe
11	H	Cl	Cl	H	H	*t*-Bu
12	H	H	SCF_3	H	H	Me
13	Cl	H	H	CF_3	H	*t*-Bu
14	H	H	OCF_2–	O-CF_2	H	Me
15	Cl	H	H	CF_3	H	Me
16	Cl	H	Cl	Cl	H	*t*-Bu
17	H	H	SCF_3	H	H	*t*-Bu
18	H	Cl	H	H	H	*t*-Bu
19	CH_3	H	Cl	H	H	*t*-Bu
20	NO_2	H	NO_2	H	H	*t*-Bu
21	H	Cl	H	Cl	H	Me
22	CF_3	H	Cl	H	H	Me
23	H	Cl	Cl	H	H	Me
24	H	Cl	H	Cl	H	OMe
25	Cl	H	Cl	Cl	H	OEt
26	Cl	CN	Cl	H	Cl	OMe
27	Cl	H	CF_3	H	H	OMe
28	CF_3	H	Cl	H	H	OMe
29	Cl	H	CF_3	H	Cl	OEt
30	Cl	H	CF_3	H	Cl	OMe
31	SO_2Et	H	H	CF_3	H	Me

Uncoupling Activity

The uncoupling activity was measured with rat liver mitochondria with succinate as respiratory substrate (*2*). pI_{50} is defined as the negative log of the concentration of an uncoupler yielding half-maximal stimulation of oxygen uptake. We measured our pK_a values in 50% ethanol, the log *P*'s in the octanol/pH 7-buffer system and R_M values on paraffin–oil-coated commercial thin-layer plates (Sil-G, Macherey u. Nagel, Düren, Germany).

and Uncoupling Activity



Mp, °C	pK_a *Lost*	*log* P	R_M	pI_{50}
151-3	6.60	4.96	−0.535	6.86
128	6.20	5.44	−0.594	7.09
223	6.80	5.03	−0.572	7.00
227-8	6.75		−0.503	7.00
143-6	6.60	5.16	−0.669	6.96
131-5	7.15	3.82	−0.476	6.63
130	6.90	4.86	−0.594	6.89
158-9	6.60	5.02	−0.656	6.89
151	7.30	4.86	−0.246	6.85
203	6.60	5.22	−0.402	6.77
216-8	7.10		−0.239	6.74
155	6.70	5.04	−0.652	6.72
163-4	6.40	5.31	−0.254	6.72
167	5.90	5.31	−0.740	6.72
114-8	5.65	5.08	−0.809	6.68
161	6.10	5.86	−0.511	6.67
99	7.45		−0.105	6.57
156-7	7.50	4.62	−0.294	6.57
173-6	7.60	4.31	+0.075	6.57
156-8	5.85		−0.819	6.57
223	6.00	4.68	−0.736	6.57
109-10	6.00		−0.634	6.54
202-3	6.25	4.56	−0.740	6.51
195-7	7.10	4.50	−0.600	6.51
133-4	7.10	5.21	−0.467	6.47
164	5.60		−1.224	6.47
137-9	7.10	4.42	−0.457	6.44
173-4	6.90	4.66	−0.652	6.44
116-7	6.25		−0.320	6.39
122	6.20	4.36	−0.830	6.37
189-90	4.65	4.22	−0.904	6.33

Table I.

No.	X_2	X_3	X_4	X_5	X_6	R
32	CF_3	H	Cl	H	H	*t*-Bu
33	H	Cl	Cl	H	H	OEt
34	NO_2	H	CF_3	H	H	OMe
35	H	CF_3	H	CF_3	H	Me
36	H	Cl	H	H	H	Me
37	Cl	H	Cl	Cl	H	OMe [b]
38	H	H	SO_2CF_3	H	H	OMe
39	SO_2Me	H	NO_2	H	H	OMe
40	H	H	Cl	H	H	Me
41	H	H	SCF_3	H	H	OMe
42	H	H	CF_3	H	H	OMe
43	Cl	H	Cl	Cl	H	Me
44	CH_3	H	Cl	H	H	Me
45	Cl	H	NO_2	H	Cl	OMe
46	H	CHF_2	Cl	H	H	OMe
47	Cl	H	H	CF_3	H	OEt
48	Cl	H	Cl	Cl	H	OMe [a]
49	Cl	H	NO_2	H	Cl	OEt
50	SO_2Et	H	H	CF_3	H	OMe
51	H	CF_3	H	H	H	OMe
52	NO_2	H	NO_2	H	H	Me
53	H	Cl	H	H	H	OEt
54	NO_2	H	NO_2	H	H	OEt
55	H	Cl	H	H	H	OMe
56	CF_3	H	H	H	H	OMe
57	Cl	H	H	H	H	Me
58	CN	H	H	CN	H	OMe
59	Cl	H	H	H	H	OEt
60	H	H	H	H	H	OMe

[a] Trans.
[b] Cis.

Regression equations with these molecular parameters containing linear, quadratic, and interaction terms are shown in Table II. Neither of these equations gave a very satisfactory correlation. One conclusion, however, can be drawn: the negative sign of the pK_{exp2} term indicates a p*K* optimum which agrees with earlier experience—*e.g.*, with phenols. The terms with the hydrophobic parameters log *P* and R_M show little significance if any. Some improvement is gained by introducing the interaction terms $pK \cdot \log P$ and $pK \cdot R_M$; however the significance of the hydrophobic terms is still low as judged by the 95% confidence intervals. Thus, the substituents in the different positions must contribute in different ways to hydrophobic bonding of the molecule to the receptor site. A necessary consequence would be to separate the substituent effects in the regression equations by replacing the molecular parameters by sub-

Continued

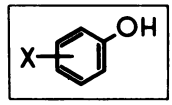
Mp, °C	pK_a *Lost*	*log* P	R_M	pI_{50}
145-7	6.20	5.67	−0.683	6.46
121-3	7.35	4.66	−0.360	6.28
172	7.20	3.44		6.27
158	5.50		−0.633	6.33
188-90	6.80	3.91	−0.612	6.25
197	6.85	5.15	−0.423	6.25
150	6.55	4.22	−0.602	6.27
254-6	5.00	3.21	−0.866	6.22
228-31	6.55	4.13	−0.648	6.22
128	7.60	4.27	−0.344	6.22
115	7.65	3.79	−0.236	6.21
138-40	5.35		−0.740	6.21
145-7	6.85	4.28	−0.492	6.19
159-60	6.00		−0.957	6.19
136	7.40	3.83	−0.378	6.19
122-4	7.40	4.45	−0.168	6.17
164	6.85		−0.481	6.15
146-7	6.00		−0.975	6.14
187-9	6.90		−1.205	6.05
103	7.60	3.78	−0.354	6.04
150-1	5.20	3.41	−1.113	5.92
144	7.80	3.94	−0.205	5.78
184-5	6.10	4.14	−0.915	5.77
125-7	7.70	3.56	−0.253	5.76
156-7	7.75	3.72	−0.308	5.73
117-9	6.75	4.01	−0.585	5.94
169	6.20	2.81	−1.014	5.72
140-2	8.50	3.38	+0.026	5.22
148	8.40	2.59	−0.159	4.96

[a] Trans.
[b] Cis.

stituents constants. In order not to increase the number of variables in the equations too much, we did this stepwise.

Substituent Parameters

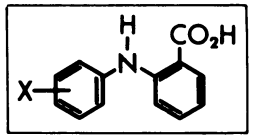
Substituent parameters offer another practical advantage. When one uses the Hansch approach to synthesize active compounds [*i.e.*, as one should call it since the Medicinal Chemistry symposium in London, April 1969 (*10*)], in a *predictive* sense, one has to rely on more or less structure-independent parameters which are accumulated in the literature or were obtained from one's own measurements. In general, the extrathermodynamic constants σ and π have proved useful in many structure–



C. Hansch, K. Kiehs, G. Lawrence
J. Am. Chem. Soc. 87, 5770 (1965)

$\log \frac{1}{c} = 6.792 - 0.491 \cdot pKa + 0.520 \cdot \pi$ — n = 14, r = 0.936, s = 0.406

$\log \frac{1}{c} = 6.607 - 0.501 \cdot pKa + 0.400 \cdot \pi$ — n = 6, r = 0.999, s = 0.039



S. Muraoka, N. Terada
Intern. Conf. on Bioenergetics
Pugnochioso, 11.–14. 9. 1970

$\log \frac{1}{c} = 2.043 + 0.777 \cdot \log P_o$ — n = 11, r = 0.723, s = 0.369

$\log \frac{1}{c} = 0.494 + 0.586 \cdot pK$ — n = 13, r = 0.808, s = 0.293

$\log \frac{1}{c} = 3.415 + 0.598 \cdot \sigma + 0.597 \cdot \pi$ — n = 11, r = 0.935, s = 0.168

Figure 3. Hansch correlations with uncouplers

Table II. Regression Equations with Molecular Parameters

Eq. No.	a_o	pK		log P		R_M		pK·logP	pK·R_M	r	s	n
		exp. 1	exp. 2	exp. 1	exp. 2	exp. 1	exp. 2					
1	−1.89 *3.45*	+3.29 *1.26*	−0.32 *0.11*							0.664	0.335	60
2	−0.05 *3.30*	+2.30 *1.28*	−0.23 *0.12*	+0.18 *0.10*						0.738	0.306	60
3	+0.71 *3.40*	+2.22 *1.26*	−0.22 *0.12*	−0.17 *0.45*	+0.05 *0.06*					0.750	0.302	60
4	−0.88 *4.30*	+2.86 *1.58*	−0.28 *0.14*			+0.04 *0.40*	+0.25 *0.37*			0.691	0.330	60
5	+5.36 *4.83*	+0.95 *1.55*	−0.16 *0.12*	−0.86 *0.69*	+0.03 *0.06*			+0.16 *0.12*		0.782	0.287	60
6	−1.60 *5.24*	+3.22 *1.72*	−0.32 *0.14*			−3.36 *4.32*	−1.42 *1.16*		0.12 *0.09*	0.832	0.255	60

$$pK_A = 7.283 - 1.838 \cdot \sigma$$

$r = 0.969$
$s = 0.16$
$n = 20$ (m, p-substituted compounds only)

R	CH_3	$C(CH_3)_3$	OCH_3	OC_2H_5
σ	0.00	−0.43	−0.60	−0.65

Figure 4. Values of pK_a in 50% ethanol/water, v/v

activity correlations despite considerable multiplicity in σ and to a lesser extent, in π.

A number of σ and π values for substituents which were present within our set of compounds were missing in the published compilations. We calculated them from the regression equations in Figures 4 and 5. An interesting feature of both equations is that the substituent R at the carbonyl group of the hydrazones contributes in a very different way from that of the substituents in the aromatic ring to p*K* as well as to log *P*. R_M values from reversed-phase thin-layer chromatography (TLC) correlated poorly with the π constants and thus cannot be used in this case to derive hydrophobic substituent parameters (*18*).

We suspect that reversed-phase TLC with these compounds is not a neat partition process although in a number of cases R_M values have proved useful (*11–13*) for correlations with biological activity. The first steps we did in separating the substituents are shown in Table III.

$$\log P_0 = 2.838 + 1.013 \cdot \Pi_{1-5} + 0.407 \cdot \Pi_R$$

$r = 0.963$
$s = 0.16$
$n = 24$

Figure 5. Partition coefficients in n-octanol/pH 7 buffer

Table III. Separation of X_1–X_5 and R

Eq. No.	a_0	Electronic Parameters		Hydrophobic Parameters		r	s	n
		$\sigma_{X\,1-5,\,R}$		$\pi_{X\,1-5,\,R}$				
		exp. 1	exp. 2	exp. 1	exp. 2			
7	+5.66 *0.19*	+0.48 *0.29*	−0.28 *0.28*	+0.31 *0.09*		0.783	0.27	60
8	+5.52 *0.29*	+0.45 *0.29*	−0.26 *0.28*	+0.50 *0.29*	−0.05 *0.07*	0.791	0.27	60

		$\sigma_{X\,1-5}$		σ_R		$\pi_{X\,1-5}$		π_R				
		exp. 1	exp. 2	exp. 1	exp. 2	exp. 1	exp. 2	exp. 1	exp. 2			
9	+5.49 *0.31*	+0.69 *0.83*	−0.28 *0.45*	+0.22 *0.45*	−0.04 *0.71*	+0.36		0.26 *0.14*		0.783	0.27	60
10	+5.41 *0.37*	+0.68 *0.83*	−0.26 *0.46*	+0.24 *0.45*	−0.01 *0.72*	+0.51 *0.46*	−0.07 *0.18*	0.26 *0.14*		0.785	0.28	60

We did not get much further than with the molecular parameters when we separated the substituents at the carbonyl group from those at the aromatic ring only. Equations 7–10 have one thing in common with Equations 1–6, the negative sign of all electronic parameter square terms. The significance of several coefficients is weak, particularly in the four-parameter equations (9 and 10). Further separation of the substituents is necessary. Some steps which gave the equations with the best fit are shown in Table IV. The equations were improved by separating the electronic parameters of the ortho substituents from all the others (Equation 11). At the same time, the hydrophobic parameters became more significant. Replacement of the ortho-σ constants by Taft's E_s values, modified according to Hansch and Kutter (*14*) gave no significant improvement. This is not surprising since there is high intercorrelation of σ-ortho and E_s within our set of substituents.

The 95% confidence intervals of the coefficients in Equations 11 and 12 indicated little significance of the electronic and steric parameters of the ortho substituents. In fact, better correlations were obtained when they were omitted and further separation of the hydrophobic constants was introduced as in Equations 13 and 14. All the parameters are highly significant except the linear π term for the ortho substituents ($\pi_{X_{1,5}}$). The correlation coefficient and standard deviation improved only slightly when a $(\pi_{X_{2\text{-}4}})^2$ term was added, as shown in Equation 15 (Figure 6).

Again, the $\pi_{X_{1,5}}$ term and the added $(\pi_{X_{2-4}})^2$ term are not significant. When we omit them, however, the correlation coefficient and standard deviation become worse.

In Figure 7 we have translated the best fit equation into the language of the chemist. The particular substituents are replaced by the increments of their electronic and hydrophobic contributions to the biological activity of the parent molecule, which has a pI_{50} of 5.81. Those acquainted with the problem of correlating structure and biological activity will agree that the Hansch approach has taught the chemist who works with biologically active molecules to think in terms of substituent parameters and coefficients rather than in terms of fluorine, chlorine, bromine, iodine, or methyl, ethyl, propyl, isopropyl, butyl, and so forth.

Some conclusions can be drawn from the best equation (Equation 15). For the meta and para substituents a σ optimum value can be calculated: $\sigma_{opt} = 0.326$, which is the order of magnitude of σ values for chlorine (*m:* 0.373, *p:* 0.227) or CF_3 (*m:* 0.430). The π^2 terms for *m,p*-substituents are not significant; therefore high lipophilicity is favorable. The same is true for the substituent at the CO group.

The role of the ortho substituents as described by our equation is entirely different. The negative sign of the linear and the positive sign of the quadratic π term suggest a minimum value for pI_{50} with $\pi = 0.18$. This means that the activity increases with more lipophilic and with more

Table IV. Separation of $X_{1,5,2,3,4}$ and R

Eq. No.	a_0	Electronic Parameters				Hydrophobic/Steric Parameters				r	s	n
		$\sigma_{X\,2-4,R}$		$\sigma_{X\,1,5}$		$\pi_{X\,1-5}$	π_R	$E_{sX\,1,5}$				
		exp. 1	exp. 2	exp. 1	exp. 2	exp. 1	exp. 1	exp. 1	exp. 2			
11	+5.79 *0.18*	+0.54 *0.19*	−0.39 *0.28*	−0.58 *0.51*	+0.60 *0.59*	+0.36 *0.10*	+0.22 *0.11*			0.871	0.22	60
12	+5.68 *0.17*	+0.53 *0.19*	0.40 *0.23*			+0.36 *0.09*	+0.22 *0.10*	−0.13 *0.13*	+0.07 *0.05*	0.877	0.21	60

		$\sigma_{X\,2-4,R}$		$\pi_{X\,2-4}$	π_R	$\pi_{X\,1-5}$					
		exp. 1	exp. 2	exp. 1	exp. 1	exp. 1	exp. 2				
13	+5.83	+0.56 *0.16*	−0.82 *0.30*	0.38 *0.08*	0.17 *0.09*	0.18 *0.12*		0,910	0.18	60	
14	+5.78	+0.51 *0.16*	−0.77 *0.29*	0.41 *0.08*	0.19 *0.08*	−0.15 *0.29*	0.37 *0.30*	0.920	0.17	60	

$$pI_{50} = 5.83 + 0.52 \cdot \boxed{\sigma_{X2-4,R}} - 0.80 \cdot \boxed{(\sigma_{X2-4,R})^2}$$
$$0.15 \quad 0.16 \qquad\qquad 0.30$$

$$+0.31 \cdot \boxed{\pi_{X2-4}} + 0.05 \boxed{(\pi_{X2-4})^2} - 0.13 \cdot \boxed{\pi_{X1,5}}$$
$$0.23 \qquad\qquad 0.10 \qquad\qquad 0.29$$

$$+0.36 \cdot \boxed{(\pi_{X1,5})^2} + 0.19 \cdot \boxed{\pi_R}$$
$$0.30 \qquad\qquad 0.08$$

$$r = 0.921$$
$$s = 0.172$$
$$n = 60$$

Figure 6. Equation 15. This equation gave the best correlation.

hydrophilic ortho substituents than, for example, OC_2H_5 ($\pi_{ortho} = 0.24$). This result is improbable. Since the $\pi_{X_{1,5}}$ term is not particularly significant, these π terms are probably not an accurate description of steric and hydrophobic effects of the ortho substituents. The minimum value is difficult to interpret. In part it may be caused by little variance in our ortho substituents. However, Equations 11 and 12 which included σ_0 or E_s parameters have also indicated minimum rather than maximum values.

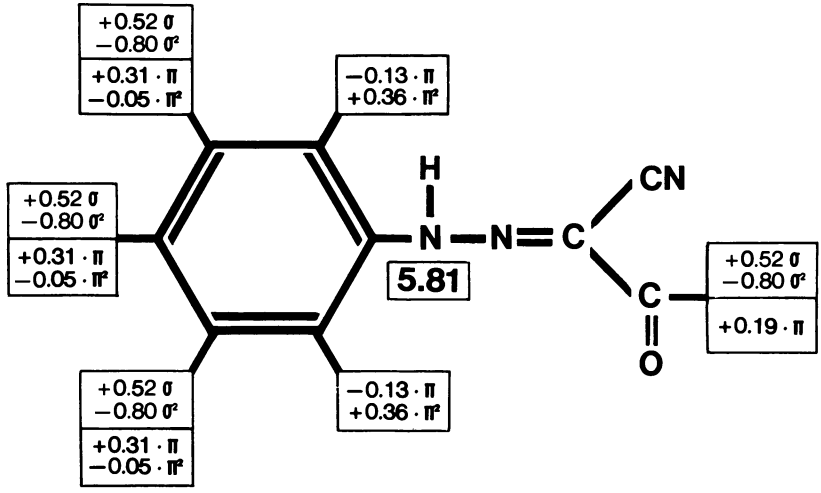


Figure 7. Contribution of substituents in different positions of the molecule to uncoupling activity

Conclusions

Although the effect of the ortho substituents needs further elucidation which requires the synthesis of new compounds, we tentatively draw one conclusion: it seems likely that steric interaction of the ortho substituents with the toxophoric NH group does not decrease the activity; on the contrary, it may enhance it.

This surprising result does not agree with any hypothesis which would involve a chemical reaction at the NH group—*e.g.* with a high energy intermediate as postulated by the Slater theory on the mechanism of oxidative phosphorylation (*15*). Conversely, hydrophobic shielding of the NH group seems advantageous. On the other hand, the optimal pK value which can be calculated from $\sigma_{opt} = 0.326$ and the equation

$$pK_a = 7.283 - 1.838 \cdot \sigma \text{ (see Figure 4)}$$

as pK_a (opt) = 4.89 (50% ethanol/water, v/v) should be precise. Hence, the capability of the uncoupler to dissociate in the biological system is essential for its mode of action. This conclusion would be in line with the chemiosmotic theory of uncoupling action by Mitchell (*16*) although this theory does not help to explain the obvious diverse contribution of the substituents in different positions of the molecule. Our correlations, though far from perfect, have clearly demonstrated that uncoupling cannot be explained simply in terms of dissociation in homogeneous phase and lipophilicity of the undissociated molecule as a whole.

A two-dimensional and perhaps stereospecific interaction of the uncoupler with its environment seems to control the activity. Generally, this would mean binding to a membrane surface. An alternative has recently been discussed by Finkelstein (*17*). In agreement with the Mitchell theory he found that the conductance of thin lipid membranes was increased by uncouplers. As a charge carrier for hydrogen ions through the membrane he suggested a dimer formed from undissociated and dissociated uncoupler molecules. Complexes of this type would be fairly lipophilic and therefore soluble in the lipid membranes. There is, however, no direct experimental evidence of their existence although they could form in solvents of low water content. One factor controlling uncoupling activity would then be the capability to form negatively charged dimers. On the other hand, monomeric uncoupler anions should also be sufficiently lipophilic when the negative charge is buried under hydrophobic substituents, and hydration is therefore reduced.

These considerations show that quantitative structure–activity correlations with uncouplers can be used as a predictive tool (*18*) not only for

the synthesizing chemist but perhaps also in a diagnostic sense for those trying to elucidate the mechanism of oxidative phosphorylation. In this respect, the present study is just a beginning, and it might be exciting to compare it with results obtained with other types of uncouplers.

Acknowledgments

We gratefully acknowledge the help of G. Schäfer, Medizinische Hochschule Hannover, who measured the biological data, and of F. Ligensa, Farbenfabriken Bayer AG, Leverkusen, who carried out the computations.

Literature Cited

1. Büchel, K. H., Korte, F., Beechey, R. B., *Angew. Chem.* (1965) **77,** 814.
2. Büchel, K. H., Schäfer, G., *Z. Naturforsch.* (1970) **25 b,** 1465.
3. Weinbach, E. C., Garbus, J., *J. Biol. Chem.* (1965) **240,** 1811.
4. Röchling, H., Büchel, K. H., *Z. Naturforsch.* (1970) **25 b,** 1103.
5. Corbett, J. R., Wright, B. J., *Pestic. Sci.* (1970) **1,** 120.
6. Büchel, K. H., Draber, W., *Progr. Photosynth. Res.* (1969) **III,** 1777.
7. Büchel, K. H., Draber, W., Trebst, A., Pistorius, E., *Z. Naturforsch.* (1966) **21 b,** 243.
8. Büchel, K. H., Draber, W., Hammann, G., Netherlands Patent **6,813,353** (March 28, 1969).
9. Heytler, P. G., *Biochemistry* (1963) **2,** 357.
10. Tute, M. S., *J. Med. Chem.* (1970) **13,** 48.
11. Boyce, C. B. C., Milborrow, B. V., *Nature* (1967) **208,** 2275.
12. Gabbott, P. A., *Progr. Photosyn. Res.* (1969) **III,** 1712.
13. Draber, W., Büchel, K. H., Dickoré, K., *Proc. Intern. Congr. Pesticide Chem., 2nd, Israel, Feb. 1971* (1972) **V,** 153.
14. Kutter, E., Hansch, C., *J. Med. Chem.* (1969) **12,** 647.
15. Slater, E. C., in "Regulation of Metabolic Processes in Mitochondria," J. M. Tager, S. Papa, E. Quagliariello, E. C. Slater, Eds., pp. 166–178, Amsterdam, 1966.
16. Mitchell, P., *Biol. Rev.* (1966) **41,** 445.
17. Finkelstein, A., *Biochem. Biophys. Acta* (1970) **205,** 1.
18. Draber, W., Büchel, K. H., Schäfer, G., *Z. Naturforsch.* (1972) **27b,** 159.

Received May 10, 1971.

Structure-Activity Relationships in Antifungal Agents: A Survey

ERIC J. LIEN

School of Pharmacy, University of Southern California, Los Angeles, Calif. 90007

CORWIN HANSCH

Department of Chemistry, Pomona College, Claremont, Calif. 91711

Using computerized regression analysis, a generalized equation and a few of its simplified forms can be used to correlate the antifungal activity of more than 560 compounds with their chemical structures. The general equation is:

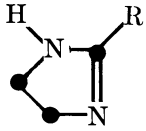
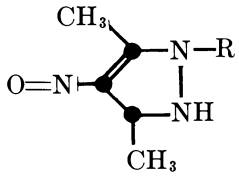
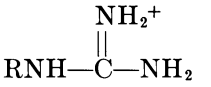
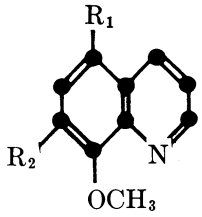
$$\log \text{activity} = \text{parabolic function of } \log P + k \text{ (electronic)} + k' \text{ (steric)} + k''$$

Hydrophobic character as measured by the octanol/water partition coefficient (log P or π) and electronic effects of substituents as measured by Hammett's σ constant appear to be the most important factors in determining the relative potency of congeneric members of drugs while the intrinsic activity is governed by the functional group(s) of the molecules. The correlations obtained should serve as useful guidelines for predicting the design of more specific new antifungal agents.

Among the 70,000 known species of fungi many are parasitic to animals and plants. Furthermore, under proper conditions almost any material is open to fungal attack if not adequately protected by a fungicide. It is not surprising, then, that thousands of compounds are synthesized and tested each year for antifungal activity. A number of antifungal agents have thus become commercially available for various purposes (*1–3*).

Table Ia. Equations Correlating Antifungal

$$\log \text{Activity} = k_1 \log P + k_2$$

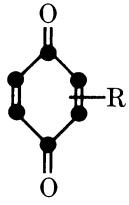
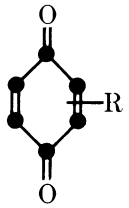
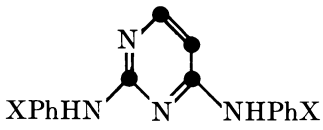
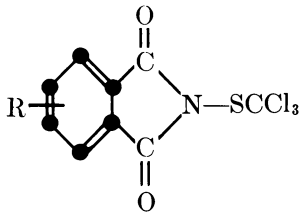
Organism	*Type of Compound*	*Action*	*Units of Activity*
S. sarcinaeforme spores		inh.	$1/ED_{50}$ $\mu M/cm^2$
S. sarcinaeforme		inh.	$1/ED_{50}$ $\mu M/cm^2$
V. inaequales spores		inh.	% inh. germination
T. mentagrophytes		inh.	p*C*
M. verrucaria	"	inh.	p*C*
C. albicans	HOC_6H_4COOR	inh.	p*C*
T. mentagrophytes	$R_1RN(CH_2)_2N(CH_3)_2$	inh	p*C*
C. albicans	$R_1R_2N(CH_2)_2N(CH_3)_2$	inh.	p*C*
A. niger	$RCOO^-$	kill	p*C*
T. interdigitale	$RCOO^-$	kill	p*C*
A. niger	$RCOO^-$	inh.	p*C*
P. cyclopium	$XPh(CH_2)_2NCS$	inh.	p*C*
A. niger	$XPhCH_2NCS$	inh.	p*C*

Activity with Physicochemical Constants

k_1	k_2	n [a]	r [b]	s [c]	*Equation*	*Ref.*
0.51	−0.94 [d]	3	0.99	0.24	1	*4*
1.65	0.06 [b]	3	0.99	0.30	2	*4*
0.15	1.51 [d]	5	0.94	0.16	3	*4*
0.81	0.90	12	0.95	0.16	4	*4*
0.79	0.93	9	0.93	0.15	5	*4*
0.70	0.95	7	0.97	0.21	6	*8*
0.53	1.37	22	0.89	0.50	7	*8*
0.34	1.74	19	0.86	0.40	8	*8*
0.67	2.08	8	0.97	0.18	9	*8*
0.76	2.43	14	0.99	0.13	10	*8*
0.55	2.66	10	0.96	0.21	11	*8*
0.46	2.79	6	0.97	0.06	12	*8*
0.55	3.28	13	0.90	0.15	13	*8*

Table Ia.

$$\log \text{Activity} = k_1 \log P + k_2$$

Organism	*Type of Compound*	*Action*	*Units of Activity*
M. fructicola		kill	p*C*
A. oleracea		kill	p*C*
C. albicans		inh.	p*C*
N. crassa		inh.	p*C*

[a] Footnotes for Tables Ia, b, c are given in Table Id (p. 176).

In spite of the tremendous amount of data reported in the literature, few generalized quantitative structure-activity correlation studies have been reported (*4–8*); that is, little investigation, using a generalized model, into correlations between the drug's molecular structure and the resultant biological activity has been done. We now report that with the use of computerized regression analysis, a generalized equation and a few of its simplified forms can be used to correlate the antifungal activity of more than 560 compounds with their chemical structures. The correlations obtained should serve as guidelines for the design of new antifungal agents.

Continued

k_1	k_2	n [a]	r [b]	s [d]	*Equation*	*Ref.*
0.88	3.53	10	0.86	0.58	14	*8*
0.73	3.74	10	0.83	0.56	15	*8*
0.50	4.15	8	0.96	0.22	16	*8*
0.55	4.37 [e]	7	0.92	0.21	17	*7*

Method

The biological data were collected from a survey of the literature up to December 1970. Although an enormous amount of work has been published, only data suitable for quantitative analysis could be considered. The physicochemical constants used in the study were log P or π, where P is the octanol/water partition coefficient of the whole molecule and π is defined as:

$$\pi = \log P_X - \log P_H$$

(P_X is the partition coefficient of a derivative and P_H is that for the parent compound.) Also used were Hammett's σ constant, Taft's polar constant, σ^*, and Taft's steric parameter, E_s. In a few examples (Equations 17, 21, 24, and 30), P values from oleyl alcohol/water have been used. In one instance (Equation 69) the chemical shift of a phenolic proton has been used for comparison with the σ constant. Where possible, the experimentally measured partition coefficients for all members of the series have been used. In other instances only one member of a set has been measured. Values for the other members were obtained by taking advantage of the additivity principles of log P and π. Details are given elsewhere (*4*, *7*, and *8*). For the new work of Table II, log P values for the parent compounds are given in the footnotes.

The "best" equations are assembled in Tables Ia, b, and c. Here we have given the equations with the maximum number of independent variables justified by the F statistic where $\alpha \leqslant 0.10$. Many sets were studied in which only poor correlations were obtained. We have not included these. Our standard for a good correlation was set at $r \geqslant 0.9$ (r = correlation coefficient). Only a few examples with r slightly below 0.9 have been included. At present we are trying to establish a basic set of equations with which others can be compared in quantitative studies. For practical work in designing new fungicides one would want to use equations having lower correlations for guidance to design new derivatives for synthesis. In Tables Ia, b, and c, n is the number of data points used in the least-squares fit of the data and s is the standard deviation from the regression. Log P_o was obtained by setting $\partial(\log \text{activity})/\partial \log P = 0$ and solving for log P. It represents the optimum lipophilic character for the given set of congeners. The figures in parentheses under this constant are the 95% confidence intervals. In some instances confidence intervals are missing because it is not possible to calculate them. An explanation of this calculation is given elsewhere (*9*).

Where possible, activity has been expressed as log $1/C$ (*i.e.*, pC) where C is the molar concentration required to cause a standard response (such as ED_{50}, MIC, or LD_{100}). In many instances the intercepts of these equations can be compared. Where activity is expressed in other units, such comparisons are not possible. In a few examples the relative value, PC', the molar phenol coefficient, has been used. In Tables IIa-d new data not previously correlated are assembled.

Results and Discussion

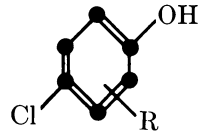
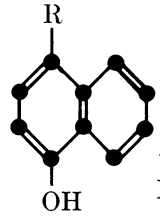
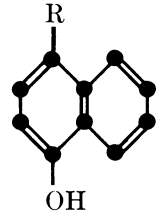
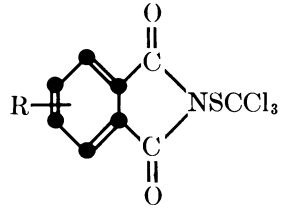
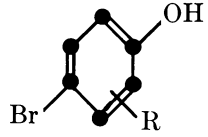
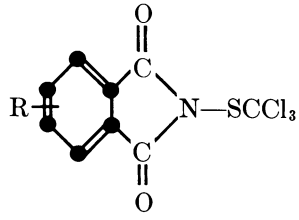
In Tables Ia and Ib we have placed all equations which are correlated by the single parameter, log P. Except for those equations in which activity could not be defined by pC (*i.e.*, log $1/C$), all equations have

been arranged by increasing value of the intercept. Since activity for these is defined as the reciprocal of the molar concentration of fungicide ($1/C$), the larger the value of the intercept, the greater the intrinsic activity of the pharmacophoric function of a given congeneric set. In comparing the intercept of these equations we are considering the case where $P = 1$ or $\log P = 0$. Comparing intercepts thus allows one to compare completely different sets of congeners acting on completely different biochemical systems under the condition where the molecules have the same lipophilic character. If comparison of two or more sets of congeners is being made on an identical test system (same organism, temperature, nutrient, and so forth), then differences in intercept can be taken as differences in what might be called the intrinsic activity of the common pharmacophoric function. Stated another way, if $\log P$ accounts for differences in activity caused by the hydrophobic character of the drugs (and this is the only variable in our equation), then other differences between sets are contained in the intercept. At our present level of refinement in extrathermodynamic correlations these might be lumped together under the common heading stereoelectronic. Fortunately, it turns out (*10*) that when the octanol/water reference system is used as a standard, the intercept for the most nonspecific kinds of biological response is $0.0 \pm .5$; that is, equations (in the single variable $\log P$) correlating simple protein denaturation, narcotic action on frog hearts, narcosis of tadpoles, and the like have intercepts near zero. This holds only for neutral molecules such as alcohols, ketones, esters, and so forth. Ionic compounds such as $RCOO^-$ and $RN^+(CH_3)_3$ deviate from this greatly. It must also be kept in mind that the value of the intercept depends on the level of response; that is, the intercept for an ED_{100} would be lower than that for an ED_{50}.

We can cautiously begin to use intercepts such as those in Tables Ia and Ib to order various functional groups in terms of their relative effects on various biochemical systems. There is an advantage in making comparisons of molecules on an isolipophilic basis early in structure-activity studies, which can be illustrated as follows. Omitting the two extreme examples (Equations 2 and 3, which are based on little data), the mean and standard deviation for the slopes of the other 15 cases in Table Ia is 0.62 ± 0.15. On this basis, molecules in different congeneric sets which differ by 3 in $\log P$ values would have differences in activity of 1.86 log units, or 70-fold. While this would indicate greatly different activity for the two isolated examples, if the intercepts for the two equations correlating the SAR for the sets are the same, little importance is to be attached to the 70-fold difference in activity. This can be achieved for any members of the sets simply by manipulation of the $\log P$ values.

Table Ib. Equations Correlating Antifungal

$$\log \text{Activity} = -k_1 (\log P)^2 + k_2 \log P + k_3$$

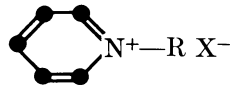
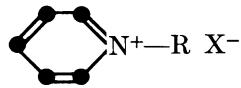
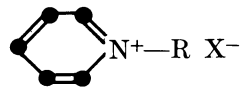
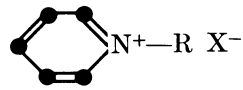
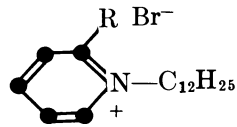
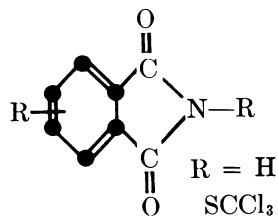
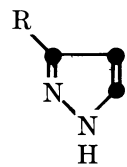
Organism	*Type of Compound*	*Action*	*Units of Activity*
T. rosaceum		inh.	*PC′*
A. niger	 with EDTA	inh. Mycelia	mm/day
A. niger		inh. Mycelia	mm/day
A. tenuis		inh.	relative
C. albicans		inh.	*PC′*
V. inaequales spores	RNH_3^+	inh.	% Germination
E. graminis		inh.	% reduction in infection

Activity with Physicochemical Constants

k_1	k_2	k_3	*log* P_o	n	r	s	*Equation*	*Ref.*
0.11	1.68	−2.68 [d]	7.3 (6.2–11)	22	0.99	0.13	18	*8*
0.14	1.36	−2.36 [d]	5.0 (4.9–5.2)	9	0.99	0.08	19	*4*
0.13	1.20	−1.84 [d]	4.7 (4.6–5.0)	11	0.96	0.11	20	*4*
0.21	1.13	−1.59 [d]	2.8 (2.3–31)	6	0.98	0.08	21	*7*
0.08	1.32	−0.89 [d]	7.8 (6.2–15)	14	0.99	0.12	22	*8*
0.18	1.38	−0.85 [d]	3.9 (3.5–5.5)	5	0.99	0.07	23	*4*
1.12	3.87	1.90 [d]	1.7 (0.7–2.0)	6	0.96	0.32	24	*7*

Table Ib.

$$\log \text{Activity} = -k_1 (\log P)^2 + k_2 \log P + k_3$$

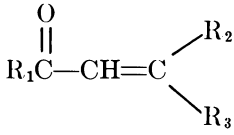
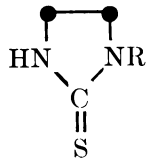
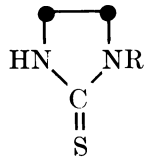
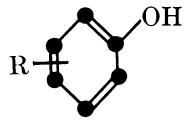
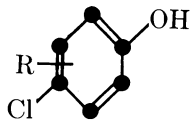
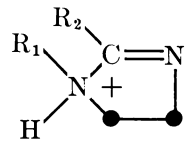
Organism	*Type of Compound*	*Action*	*Units of Activity*
M. fructicola		kill	$1/\mu M/cm^2$
P. infestans		kill	$1/\mu M/cm^2$
A. oleracea		kill	$1/\mu M/cm^2$
V. inaequales		kill	$1/\mu M/cm^2$
V. inaequales		kill	$1/\mu M/cm^2$
S. sarcinaeforme		inh.	$1/ED_{50}$ mole/cm²
T. interdigitale		inh.	p*C*
L. lepideus	ROH	inh.	p*C*

Continued

k_1	k_2	k_3	*log* P_o	n	r	s	*Equation*	*Ref.*
0.16	0.78	2.38 [d]	2.5 (2.2–3.0)	7	0.99	0.12	25	*8*
0.13	0.65	2.55 [d]	2.4 (2.0–3.2)	8	0.99	0.16	26	*8*
0.17	0.72	2.58 [d]	2.2 (1.6–3.6)	7	0.97	0.30	27	*8*
0.16	0.64	3.07 [d]	2.0 (1.8–2.3)	8	0.99	0.10	28	*8*
0.08	0.59	3.10 [d]	3.6 (3.1–11)	6	0.93	0.10	29	*8*
0.44	2.42	5.33 [d, e]	2.8 (2.4–4.0)	7	0.99	0.16	30	*7*
0.65	5.98	−9.33	4.6 (4.3–5.4)	6	0.95	0.23	31	*8*
0.60	4.27	−4.09	3.6	5	0.88	0.53	32	*8*

Table Ib.

$$\log \text{Activity} = -k_1 (\log P)^2 + k_2 \log P + k_3$$

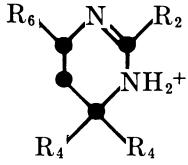
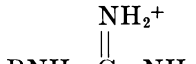
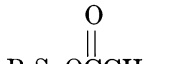
Organism	*Type of Compound*	*Action*	*Units of Activity*
A. niger		inh.	p*C*
A. solani		inh.	p*C*
C. albicans	ArNCS	inh.	p*C*
M. fructicola		inh.	p*C*
A. niger		inh.	p*C*
T. gypseum		inh.	p*C*
G. cingulata		inh.	p*C*
A. solani	"	inh.	p*C*
S. sarcinaeforme	"	inh.	p*C*
M. fructicola	"	inh.	p*C*

Continued

k_1	k_2	k_3	*log* P_o	n	r	s	*Equation*	*Ref.*
0.13	1.38	−1.60	5.1	19	0.87	0.41	33	*8*
0.34	2.18	−0.29	3.2 (2.8–9.8)	6	0.97	0.15	34	*8*
0.19	1.91	0.56	5.0 (4.6–6.5)	10	0.94	0.10	35	*8*
0.14	1.40	0.66	5.0	5	0.99	0.05	36	*8*
0.11	1.17	0.83	5.5 (5.1–6.4)	26	0.93	0.20	37	*8*
0.10	1.32	1.18	6.4	11	0.92	0.34	38	*8*
0.07	0.97	1.44	7.0 (6.0–8.7)	15	0.88	0.50	39	*8*
0.08	0.94	1.89	6.2 (5.6–6.9)	15	0.91	0.36	40	*8*
0.07	0.90	2.09	6.1 (5.6–6.8)	15	0.91	0.34	41	*8*
0.08	1.07	2.18	6.1 (5.7–6.7)	14	0.95	0.32	42	*8*

Table Ib.

$$\log \text{Activity} = -k_1 (\log P)^2 + k_2 \log P + k_3$$

Organism	*Type of Compound*	*Action*	*Units of Activity*
C. albicans	$RCOO^-$	inh.	p*C*
L. lepideus	RNH_3^+	inh.	p*C*
A. circinaus spores		inh.	p*C*
M. fructicola	"	inh.	p*C*
A. niger	$PhCH_2N^+(CH_3)_2\ Cl^-$	inh.	p*C*
T. mentagrophytes	"	kill	p*C*
T. mentagrophytes	"	inh.	p*C*
C. albicans	"	inh.	p*C*
C. albicans	"	kill	p*C*
P. omnivorum	$RCOO^-$	inh.	p*C*
T. interdigitale	"	inh.	p*C*
T. purpurpeum	"	inh.	p*C*
T. gypseum	"	inh.	p*C*
M. fructicola		inh.	p*C*
S. pastorianus spores	"	inh.	p*C*
A. niger		inh.	p*C*
R. nigricans	"	inh.	p*C*

Continued

k_1	k_2	k_3	*log* P_o	n	r	s	*Equation*	*Ref.*
0.64	−1.54	2.15	−1.2 (−1.3–1.0)	6	0.99	0.09	43	*8*
0.12	0.42	2.38	1.7	6	0.90	0.28	44	*8*
0.05	0.67	2.69	7.3	12	0.90	0.65	45	*4*
0.05	0.73	2.70	7.7	11	0.90	0.70	46	*4*
0.24	1.04	3.11	2.2 (2.0–2.5)	11	0.95	0.22	47	*8*
0.16	1.06	3.23	3.2 (2.9–3.9)	11	0.98	0.17	48	*8*
0.16	1.06	3.29	3.3 (3.0–3.7)	11	0.99	0.12	49	*8*
0.26	1.36	3.24	2.6 (2.4–2.8)	11	0.98	0.20	50	*8*
0.30	2.04	3.25	2.2 (2.1–2.5)	11	0.96	0.24	51	*8*
0.03	−0.08	3.35	−0.2 (−1–10)	15	0.85	0.35	52	*8*
0.10	0.33	3.50	1.6 (0.7–5.0)	14	0.97	0.22	53	*8*
0.06	0.46	3.75	4.2 (2.3–11)	15	0.99	0.17	54	*8*
0.12	0.24	4.25	1 (0.6–2.0)	15	0.88	0.21	55	*8*
0.13	0.52	4.84	2.0 (1.8–2.2)	6	0.99	0.03	56	*8*
0.18	0.71	4.88	1.9 (1.5–2.2)	6	0.99	0.07	57	*8*
0.14	0.86	4.48	3.0[g] (2.8–3.2)	6	0.99	0.10	58	*4*
0.15	0.76	4.58	2.5[g] (1–4.5)	6	0.92	0.57	59	*4*

Moreover, having a fair idea of log P_0 (the ideal lipophilic character for a given set acting on a given system), one can calculate rough limits on the amount of activity increase one can ultimately obtain for a given set simply by increasing the partition coefficient in the synthesis of new congeners. Comparison of isolipophilic molecules romoves one complicating factor in unraveling structure-activity relationships. To put it in practical terms, if one has two entirely different series of drugs requiring different numbers of parameters to give good correlations, one should choose one drug from each series having approximately the same log P value and compare their activities (log $1/C$). This comparison may provide more useful information for further drug design than comparing, say, the MIC of the most potent member of each series.

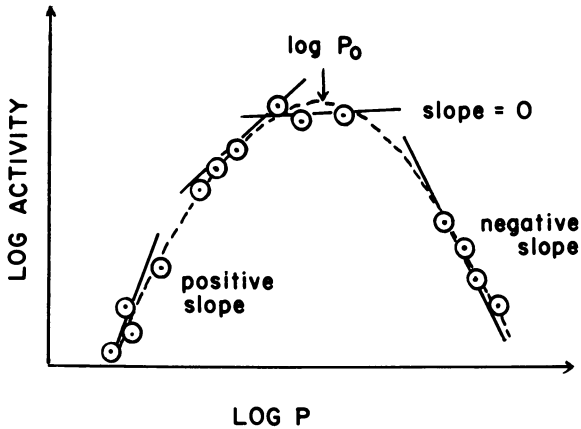


Figure 1. Dependence of log activity on log P. *Different segments of the same parabola may give straight lines with slopes ranging from positive to zero to negative.*

When making comparisons with linear equations such as those in Table Ia, one must proceed with caution. Since the general relationship between pC and log P is parabolic, the linear relation is only part of the larger picture. As Figure 1 shows, the slope and intercept for such an equation will depend strongly on the part of the parabola on which one is working. Thus, for valid comparisons the set of congeners must be suboptimal with respect to log P_0 (the apex of the parabola), and reasonable variation must be present in the log P values of the members of the sets. One must also keep in mind that although electronic and steric terms are not present in the equations of Tables Ia and Ib, this does not mean that such effects would not be important in these series. Often the

set of congeners was not well chosen so that insufficient variation is present in these properties for their weight to be assessed.

The most obvious generalization about the equations of Table Ia is that since all are linear in log P, more active members of each set could be made simply by increasing lipophilic character. However, in undertaking such a study it would be most desirable to select substituents so that the effects of electronic and steric factors could be evaluated simultaneously.

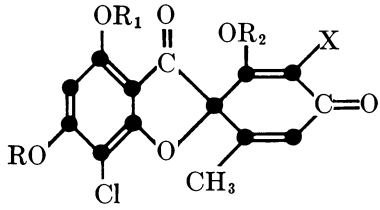
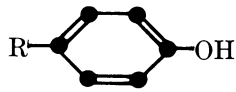
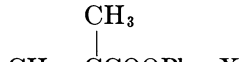
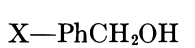
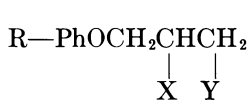
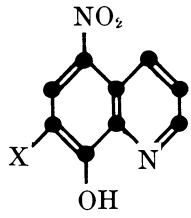
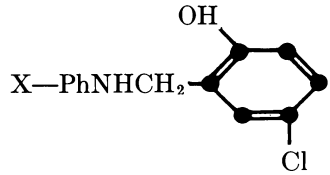
The intercept of Table Ia indicate that only the last five or six sets are active enough for further study. How much activity one could expect to gain could be estimated by noting the difference between the most active compound tested in a given set and the probable log P_0 value for the set. The log P_0 would have to be estimated by considering similar compounds in Table Ib. Most neutral sets have log P_0 in the range 5–6 while most ionic sets have log P_0 between 2 and 3. One obtains similar equations for the action of such charged molecules on red blood cells and bacteria (*10*). It would seem that these molecules are producing their toxic effects *via* membrane perturbation (*8*).

In the equations correlating sets of compounds only partially ionized under the experimental conditions (*e.g.*, RCOOH and RNH_2), one is faced with the uncertainty of which log P should be used. The choice, of course, makes a large difference in the values of the intercept and log P_0. For example, for the carboxyl function, π for COO^- is 4.1 units lower than that for the neutral COOH. While this is disadvantageous from the point of view of the penetration and binding of acids in biochemical processes, it is offset in membrane perturbation by the much greater activity of the ion (*8*, *10*). Even at a pH as low as 5, most acids would be 50% or more ionized. At physiological pH, little of the neutral form is present. For these reasons we have used log P values for the ionic forms where possible in this study. There is of course no ambiguity in the case of ionization for the quaternary ammonium compounds or the guanidines. The fact that the intercepts for equations correlating sets of $RCOO^-$ fall in the same regions as these totally ionized congeners lends support to the choice of log P for $RCOO^-$.

For the alkylguanidines, Brown and Sisler (*14*) postulated that the alteration of permeability of the cell, with the resulting loss of vital cellular constituents, blocking of anionic sites, and inactivation of vital enzymes, was responsible for the fungitoxicity. Interestingly, the correlations obtained for some of the amines and the quaternary ammonium compounds indicate activities comparable with those of the alkylguanidines. When the nonpolar side chains of these compounds are long enough, they may cause similar perturbations on the lipoprotein membranes by relatively nonspecific hydrophobic interactions and ionic forces.

Table Ic. Equations Correlating Antifungal

$$\log \text{Activity} = -k_1 (\log P)^2 + k_2 \log P + k_3 \sigma + k_4$$

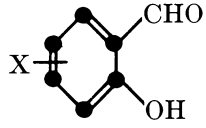
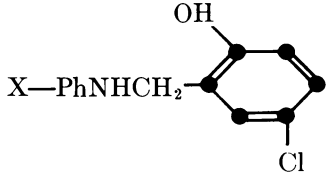
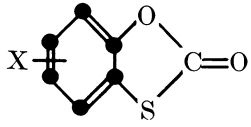
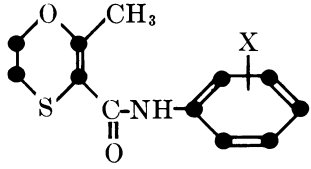
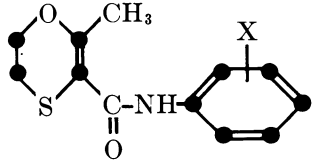
Organism	*Type of Compound*	*Action*	*Units of Activity*	k_1
B. alli		curling of hyphac	relative	0.0
A. niger		inh.	p*C*	0.19
H. anomala		inh.	p*C*	0.10
A. niger + 3 molds		inh.	p*C*	0.00
T. mentagrophytes + 3 molds		inh.	p*C*	0.00
A. niger		inh.	p*C*	0.00
M. verrucaria	"	kill	p*C*	0.00
C. albicans		inh.	p*C*	0.00

Activity with Physicochemical Constants

k_2	k_3	k_4	*log* P_o	n	r	s	*Equation*	*Ref.*
0.56	2.19	−1.32 [d]	—	22	0.88	0.25	60	*8*
1.86	0.63	−0.09	4.9 (4.3–6.3)	18	0.98	0.16	61	*8*
1.23	−0.88	0.88	6.0 (4.8–7.3)	10	0.96	0.07	62	*8*
0.58	0.45	1.12	—	19	0.95	0.24	63	*8*
0.69	0.43	1.21	—	26	0.91	0.22	64	*8*
0.00	3.51	1.52	—	4	0.996	0.03	65	*11*
0.00	4.22	1.76	—	4	0.99	0.05	66	*11*
0.27	0.96	2.43	—	8	0.98	0.11	67	*8*

Table Ic.

$$\log \text{Activity} = -k_1 (\log P)^2 + k_2 \log P + k_3 \sigma + k_4$$

Organism	*Type of Compound*	*Action*	*Units of Activity*	k_1
S. cerevisiae		inh.	p*C*	0.00
S. cerevisiae	"	inh.	p*C*	0.00
T. interdigitale		inh.	p*C*	0.00
M. audouini	"	inh.	p*C*	0.00
T. mentagrophytes		inh.	p*C*	1.09
U. maydis		inh.	p*C*	0.00
D. palmatus	"	inh.	p*C*	0.00
R. solani		inh.	p*C*	0.00

[a] *See* Table Id (p. 176) for footnotes.

Continued

k_2	k_3	k_4	*log* P_o	n	r	s	*Equation*	*Ref.*
0.56	0.41	2.72	—	14	0.89	0.35	68	*12*
0.77	0.22[f]	0.89[d]	—	14	0.86	0.40	69	*12*
0.34	0.57	3.03	—	8	0.91	0.24	70	*8*
0.24	0.74	3.54	—	8	0.98	0.10	71	*8*
1.36	−1.70	3.71[d,h]	.64 (0.4–2.2)	9	0.89	0.39	72	*4*
0.00	−1.86	5.76	—	9	0.89	0.37	73	*13*
0.00	−1.94	5.91	—	9	0.89	0.38	74	*13*
−1.58	−2.64	10.16	—	9	0.94	0.33	75	*13*

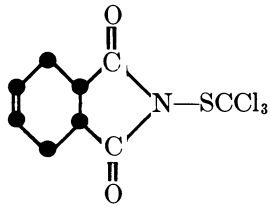
Table Id. Equations Correlating Antifungal Activity with Physicochemical Constants

Organism	*Type of Compound*	*Action*	n	r	s	*log* P_o	*Equation*	*Ref.*
A. niger	$Br—CH_2\overset{\overset{O}{\|\|}}{C}—NHR$	inh.						
	$pC = -0.26(\log P)^2 + 1.59\log P + 2.06\sigma^* + 0.83E_s + 3.20$		15	0.98	0.21	3.0 (2.8–3.2)	76	*8*
T. viride	$Br—CH_2\overset{\overset{O}{\|\|}}{C}—NHR$	inh.						
	$pC = 0.37\log P + 0.51E_s + 3.37$		10	0.95	0.20	—	77	*8*
B. cinerea	$RR'N\overset{\overset{S}{\|\|}}{C}—S^-Na^+$	inh.						
	$pC = -0.28\pi^2 - 0.21\pi - 1.53\sigma^* + 5.06$		9	0.92	0.28	−0.4	78	*8*

[a] Number of data points used in deriving equation.
[b] Correlation coefficient.
[c] Standard deviation.
[d] These intercepts not comparable with those in which data is expressed in units of log $1/C$ where C is the applied molar concentration.
[e] These equations based on oleyl alcohol/water partition coefficients. These are related to the octanol/water system by: $\log P_{\text{oley alc.}} = 0.994 \log P_{\text{octanol}} - 0.54$. Hence this intercept is about 0.5 out of line with others in the table.
[f] Instead of σ, the chemical shift of the OH was used as an index of the electronic effect of the substituent. Units are in cps $\times 10^{-2}$.
[g] Based on the experimentally measured log P of tripropyltin acetate = 1.67.
[h] Since the log P of the parent molecule is not available, the π value of the substituent X is used in the correlation.

Polyenic antibiotics such as filipin, nystatin, and amphotericin B, possessing nonpolar and polar moieties, may operate in a similar fashion. However, these compounds appear to operate *via* membrane sterols. The specific interaction with membrane sterol appears to be responsible for the selective toxicity against fungi, not bacteria, since bacterial membrane does not contain sterols. The membrane perturbation may lead to leakage of various nutrients (*15*), as well as to alteration of the activity of the enzymes incorporated into the membrane structure. Polyenic antibiotics may also cause hemolysis (*15*). We have recently pointed out the similar structure-activity correlations for antifungal agents and agents causing hemolysis (*8*).

An interesting feature of the numerical approach to structure-activity correlations has been pointed out (*7*) in the case of the phthalimide derivatives. The commercial fungicide Captan has the structure:



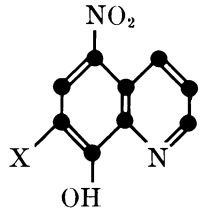
Various workers have attached special toxiphoric properties to the $SCCl_3$ function. The fact that the derivatives having either H or $SCCl_3$ attached to N fit the same equation (Equation 30) indicates that it is primarily the lipophilic character of the $SCCl_3$ function which is important. It would be most interesting to make a set of derivatives of Captan in which the $SCCl_3$ function was replaced by groups having considerable variation in electron-withdrawing properties as well as lipophilic character. If the toxiphoric action of the imide involves the cleavage of the N–$SCCl_3$ bond or the opening of the imide ring, the electronic character of the substituent on N could play a strong role. Since Captan has close to the ideal lipophilic character (*7*), replacement of the $SCCl_3$ function to study electronic effects of other substituents would have to be done carefully so that activity would not be lost by moving very far in either direction from log P_0.

Captan also has a pronounced effect on membranes. Using the electron microscope, Richmond and co-workers showed (*16*) that ED_{50} doses of Captan gave convoluted forms of the nuclear membrane to dormant conidia of *Neurospora crassa* and caused almost complete loss of intracellular fine structure after Captan-treated spores were incubated.

In the examples where there are reasonable confidence intervals on log P_0 in Table I, neutral molecules usually have log P_0 in the range 5–6. Most of the ionic sets have log P_0 in the range 2–3. Although these are crude figures on relatively few sets of congeners, they do constitute useful reference points for the synthesis of new classes of fungicides. For example, if one were working with a new neutral heterocycle, one would check early in the investigation on derivatives with log P values in the range 3.5–6.5 while if the heterocycle were charged, one would study molecules with log P in the range 1–4.

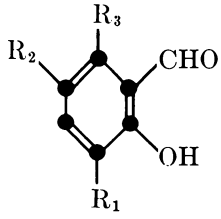
Table Ic lists equations in which an electronic term is significant. Since in all of these examples the same electronic term (σ) has been used, the intercepts are comparable. Again the intercept contains information on the stereoelectronic character of the essential pharmacophoric function. The σ parameter simply accounts for the electronic perturbation of various substituents on this intrinsic activity. Some of these equations contain a $(\log P)^2$ term, and some do not. Those equations lacking this term indicate sets comprised of molecules with suboptimal lipophilic character. More active members could be made by adding more lipophilic substituents. In almost all of these sets it should be possible to make more active derivatives by the proper combination of substituents, bearing in mind that log P_0 must first be found and then not exceeded while the electron-releasing or withdrawing (as the sign of ρ calls for) effect of the substituents is maximized.

Table IIa. Antifungal Data and the Physicochemical Constants Used in Deriving Equations



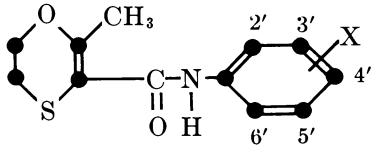
log 1/C [a]				
A. niger	M. verrucaria	*log P* [b]	σ	*X*
Inh.	*Kill.*			
3.43	4.02	2.77	0.54 [c]	F
3.89	4.66	3.21	0.68 [d]	Cl
4.01	4.72	3.41	0.70 [d]	Br
3.72	4.36	3.71	0.63 [c]	I

Table IIb. Antifungal Data and the Physicochemical Constants Used in Deriving Equations



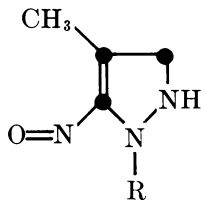
S. cerevisiae *Inh. log 1/C* [e]	*log* P[f]	σ	(*cps* × 10^{-2})	R_1	R_2	R_3
6.19	4.45	1.36	7.08	I	I	H
5.55	3.83	1.40	6.94	Br	Br	H
5.35	3.43	1.36	6.85	Cl	Cl	H
5.20	3.00	0.63	7.06	I	H	H
4.93	3.07	1.92	6.73	NO_2	Cl	H
4.85	2.94	0.70	6.58	H	Br	H
4.74	3.26	0.63	6.55	H	I	H
4.55	3.24	0.55	6.70	CH_3	Cl	H
4.47	2.74	0.68	6.58	H	Cl	H
4.11	2.29	−0.13	6.41	H	CH_3	H
3.67	1.81	0.00	6.50	H	H	H
3.63	2.79	−0.26	6.65	CH_3	CH_3	H
5.36	4.47	1.73	7.45	Cl	Cl	Cl
5.22	3.43	1.05	7.41	Cl	H	Cl

Table IIc. Antifungal Data and the Physicochemical Constants Used in Deriving Equations



log 1/C[g] U. maydis *Inh.*	D. palmatus *Inh.*	R. solani *Inh.*	*log* P[h]	σ	*X*
6.00	6.15	6.40	2.13	0.00	H
5.92	6.22	6.44	2.81	−.14[i]	2′-Me
6.30	6.30	6.52	2.64	−.07	3′-Me
4.80	5.15	5.32	2.72	0.21[i]	2′-Cl
5.38	5.58	4.48	2.89	0.37	3′-Cl
4.84	4.70	4.66	2.83	0.23	4′-Cl
4.27	4.60	5.32	1.90	0.76[i]	2′-NO_2
4.70	4.64	4.68	2.24	0.71	3′-NO_2
4.30	4.30	4.55	2.37	0.78	4′-NO_2

Table IId. Antifungal Data and the Physicochemical Constants Used in Deriving Equations

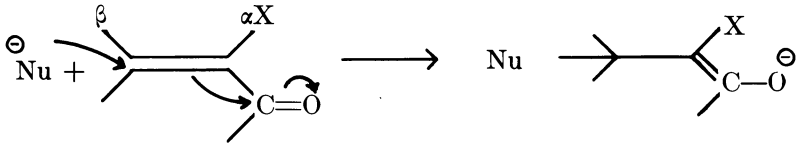


log 1/ED$_{50}$ μM/*cm*^{2} *Inh.*[j]	*log* P[k]	*R*
0.13	0.15	H
0.37	0.65	CH_3
3.77	2.28	Ph

[a] From Ref. *11*, except data for X = F from H. Gershon, private communication. Value in Ref. *11* is from 5-fluoro-7-nitro-8-quinolinol.
[b] Calculated value using log *P* = 2.02 for 8-hydroxyquinoline.
[c] From I. Biggs and R. A. Robinson, *J. Chem. Soc.* **1961,** 388.
[d] From G. B. Barlin and D. D. Perrin, *Quart. Rev.* (1966) **20,** 75.
[e] From Ref. *12*.
[f] Based on experimental log *P* of 1.81 for salicylaldehyde.
[g] From Ref. *13*.
[h] Based on experimental log *P* of 2.13 parent compound.
[i] From A. C. Farthing and B. Nam, "Steric Effects in Conjugated Systems," p. 131, Academic, New York, 1958.
[j] From S. Rich and J. G. Horsfall, *Phytopath.* (1952) **42,** 457.
[k] Based on experimental value of 2.28 for derivative where R = phenyl.

Equations 65 and 66 are unusual in that addition of a term in π does not result in an improved correlation. Only four data points are involved. Most likely these equations are in part fortuitous, and a larger selection of substituents would reveal a role for π and, possibly, E_s.

An especially interesting example is Equation 60 which correlates griseofulvin derivatives. In this equation the large coefficient with σ indicates that the electron-withdrawing group (X) promotes activity. This may be the result of the fact that griseofulvin analogs are α-β-unsaturated ketones, and nucleophilic attack at the β-position may be involved for the antifungal activity against the growing cells:



Electron-withdrawing group X would make the α-carbon more positive and therefore facilitate the reaction.

Equations 68 and 69 make an interesting comparison. Equation 68 has the electronic parameter (σ) while in Equation 69 the chemical shift of the phenolic proton has been used as suggested graphically by the original author. Although the Hammett constants do give slightly better results, this work does indicate that the chemical shift may be a useful parameter for biochemical structure-activity work. This less satisfactory result with δ is probably the result of the fact that the chemical shift is affected not only by the electronic effect but also by the magnetic effect of the substituents. Probably, the electronic component of the substituent effect is important in biochemical processes.

For the Carboxin derivatives (Equations 73–75), the hydrophobic character of the substituent has a negative role (Equation 75) or is lacking it completely. The reason for this may be that for the set of drugs tested there is little variation in π for the substituents. Log P for the parent compound may be near log P_0 (near the apex in Figure 1). It would be wise to test derivatives in which there is more variation in log P before attempting to make a decision.

Miscellaneous equations are contained in Table Id. Equation 76 is interesting in that steric, electronic, and lipophilic factors are all involved in the α-bromoacetamides. The positive dependence on both σ^* and E_s indicates that small electron-withdrawing groups enhance activity. The requirement of a large E_s value (or small size) for high activity in both Equations 76 and 77 reflects the fact that nucleophilic replacement of the α-bromo atom may be a critical step in the toxiphoric action. Electron-

releasing groups (negative σ^*) would make the α-carbon less electron deficient and retard the activity. Equation 77 lacks the σ^* terms and the $(\log P)^2$ term. The smaller number of derivatives in this set contained insufficient spread in character to evaluate the dependence on these parameters.

In formulating Equation 78, we have had to use π rather than log P values. The negative π_0 value indicates an entirely different mechanism of action as compared with the other sets of Table I.

Indeed, the connection between the chelating power of dimethyldithiocarbamic acid with copper and fungicidal activity has been well documented (*17*). The negative π_0 value suggests that the binding site for the nitrogen substituent must be in a hydrophilic region. In another study, Weuffen (*18*) reported the antifungal activity of $PHCH_2CH_2NHCSSR$ (with R ranging from methyl to amyl) to be practically independent of R.

During this study we found that some investigators reported their data in terms of percent inhibition at three different concentrations (*19, 20*). No significant correlations could be obtained from these data. Another group used the so-called relative-fungistatic activity (*21, 22*) based on the average activity against several different organisms. Although this might be a useful indication of the antifungal spectrum, it is not suitable for quantitative structure-activity correlation. Quantitative structure-activity work would certainly be advanced if all biological activities could be expressed in terms of ED_{50} or LD_{50} on a molar basis. Such values which normally fall on the linear portion of the sigmoid dose-response curve are much more comparable than ED_{100} or LD_{100} figures.

It is gratifying to see that structure and activity for so many different compounds acting against a variety of fungi can be pulled together rationally using three physicochemical parameters. Even though many of the congeneric sets were poorly designed so that great variance is not present in steric, electronic, and hydrophobic effects of the substituents, enough is present in many examples to illustrate how proper manipulation of these properties of substituents can result in the design of more active fungicides. As more such equations are generated, we shall be better equipped to design the most active members of a set based on new pharmacophoric functions. Also, as more equations are formulated on other organisms, we shall be in a much better position to design more specific antifungal agents.

Acknowledgment

We are indebted to Donald K. Nikaitani for measuring a number of the partition coefficients used in this work. We wish to thank J. G.

Horsfall of the Connecticut Agricultural Experimental Station for a sample of 1-phenyl-3,5-dimethyl-4-nitrosopyrazole, L. V. Edginton of the University of Guelph, Ontario, Canada, for a sample of Carboxin, and G. M. van der Want of the University of Utrecht, The Netherlands, for a sample of tripropyltin acetate.

Literature Cited

1. Horsfall, J. G., "Principles of Fungicidal Action," Chronica Botanica Co., Waltham, Mass., 1956.
2. Torgeson, D. C., "Fungicides," Vol. 1, Academic, New York, 1967.
3. Torgeson, D. C., "Fungicides," Vol. 2, Academic, New York, 1969.
4. Lien, E. J., "The Role of Hydrophobic Bonding in Fungitoxicity," Symp. Fungitoxal Action, Annu. Meet. Amer. Phytopathological Soc., Spokane, Wash., Aug. 19, 1969.
5. Turner, N. J., "The Role of Electronic and Steric Effects in Fungitoxicity," Symp. Fungitoxal Action, Annu. Meet. Amer. Phytopathological Soc., Spokane, Wash., Aug. 19, 1969.
6. Siegel, M. R., "Sites of Fungitoxic Action," Symp. Fungitoxic Action, 61st Annu. Meet. Amer. Phytopathological Soc., Spokane, Wash., Aug. 19, 1969.
7. Lien, E. J., *J. Agr. Food Chem.* (1969) **17,** 1265.
8. Hansch, C., Lien, E. J., *J. Med. Chem.* (1971) **14,** 653.
9. Hansch, C., Steward, A. R., Anderson, S. M., Bentley, D., *J. Med. Chem.* (1968) **11,** 1.
10. Hansch, C., Glave, W. R., *Mol. Pharmacol.* (1971) **7,** 337.
11. Gershon, H., McNeil, M. W., Hinds, Y., *J. Med. Chem.* (1969) **12,** 1115.
12. Cronenberger, L., Dolfin, B., Pacheco, H., *C. R. Acad. Sci. Paris, Ser. D.* (1969) **269,** 1334.
13. Snel, M., Von Schmeling, B., Edginton, L. V., *Phytopath.* (1970) **60,** 1164.
14. Brown, I. F., Sisler, H. D., *Phytopath.* (1960) **50,** 830.
15. Lampen, J. O., *Am. J. Clinic. Pathology* (1969) **52,** 138.
16. Richmond, D. V., Somers, E., Millington, P. F., *Ann. Appl. Biol.* (1967) **59,** 233.
17. Albert, A., "Selective Toxicity," 3rd ed., p. 259, Wiley, New York, 1965.
18. Weuffen, W., *Pharmazie* (1966) **21,** 686.
19. Lobov, V. P., Stopkan, V. V., Cherepenko, T. I., Dubenko, G. D., Peretyazhko, N. Z., *Microbiology* (1966) **35,** 277.
20. Khaskin, I. G., Shomova, E. A., Stolper, A. L., *Microbiology* (1966) **36,** 853.
21. Schönenberger, H., Bastung, T., Adam, D., *Arzneim.-Forch.* (1960) **19,** 1082.
22. Schönenberger, H., Bastung, T., *Arzneim.-Forsch.* (1970) **20,** 386.

RECEIVED August 20, 1971.

12

Structure–Activity Correlations for Meta- and Para-Substituted Trifluoromethanesulfonanilide Pre-Emergence Herbicides

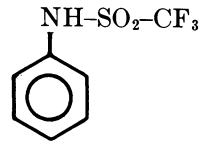
ANTHONY F. YAPEL, JR.

Central Research Laboratories, 3M Co., St. Paul, Minn. 55133

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

The Hansch equation was used to correlate pre-emergence herbicidal activity with structure for 15 meta- and para-substituted trifluoromethanesulfonanilides (TFMS). Dose–response data were collected for two grasses (Foxtail, Cheat Grass) and a broadleaf (Wild Mustard). The most significant correlations result when the herbicides are separated into meta- and para-substituted compounds and fitted separately. Each herbicidal class exhibits a unique optimum substituent π value in its action on each weed type. Substituents having large, positive Hammett sigma constants enhance herbicidal action whether substituted meta or para to the parent group. The effects of the surfactant Tween 80 on the partitioning characteristics and pre-emergence activity of the TFMS compounds were examined. Hansch relationships which quantitate these effects on herbicidal activity were derived.

Almost 15 years have elapsed since the initial preparations of trifluoromethanesulfonanilide (TFMS) and other related perfluoroalkanesulfonanilides were reported by Brice and Trott (*1*) and Burdon *et al.* (*2*).



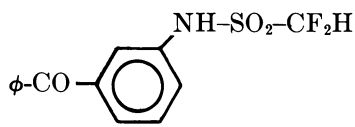
Trifluoromethanesulfonanilide (TFMS)

I

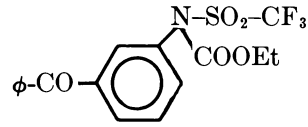
In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Since that time, mono- and difluoromethanesulfonanilides have also been synthesized (*3*), but until recently the biological activity of fluorinated TFMS compounds had not been investigated.

In early 1970, Harrington *et al.* (*4*) reported that a broad class of fluoroalkanesulfonanilides and *N*-acyl substituted fluoroalkanesulfonanilides exhibited general anti-inflammatory characteristics. Their biological screening studies indicated that two members of a series of 3-benzoyl fluoroalkanesulfonanilides [3′-benzoyl-1,1-difluoromethanesulfonanilide (II) and ethyl-*m*-benzoyl-*N*-(trifluoromethanesulfonyl)carbanilate (III)] were particularly active:



II



III

The low occurrence of undesirable side effects in experiments on common laboratory test animals (rats, mice, guinea pigs, etc.) coupled with their high biological activity made II and III especially attractive anti-inflammatory agents.

In late 1970, Trepka, Harrington, Robertson, and Waddington (*5*) reported that a number of substituted trifluoromethanesulfonanilides based on I above exhibited pre- and post-emergence herbicidal properties. Herbicidal activity was closely related to the degree of fluorination of the alkylsulfonamide side chain, with the trifluoromethanesulfonamido group shown in I being the most potent of a number of partially and completely fluorinated alkylsulfonamido parent groups examined. As shown in Table I, mono- and disubstitution in the aromatic ring of TFMS (I) significantly influenced overall pre- and post-emergence TFMS activity as well as herbicidal selectivity toward a variety of grasses, broadleaf weeds, and crops. The effect of ring substitution on herbicidal selectivity is shown by the asterisk (*)-marked TFMS monosubstituted derivatives in Table I. 3-Cl, 4-Cl, or 3-NO_2 substitution greatly enhanced the broadleaf selectivity of the TFMS parent compound while essentially eliminating herbicidal potency against grasses. Contrastingly, methylsulfonyl ($-SO_2-CH_3$) substitution in the para position eliminated the broadleaf herbicidal activity of TFMS while greatly enhancing its pre-emergence activity against grasses. Although the 2-Cl and 4-F monosubstituted series members were highly active, they did not exhibit species selectivity in their mode of herbicidal action. The 2,4-dichloro– and 2,4-difluoro–TFMS derivatives similarly failed to exhibit specificity toward either the grasses or broadleaf weeds but were highly active against all weed species

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

tested. Crop selectivity was also noted for many TFMS series members, with the 2,4-difluoro and 2,4-dichloro derivatives being particularly selective in removing grasses and broadleaf weeds from corn and soybeans (5).

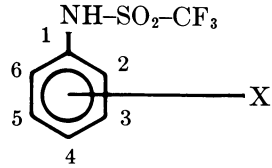
Although structure–activity studies were carried out on the substituted trifluoromethanesulfonanilides, Trepka and co-workers (5) noted that no single property of the TFMS series members tested could satisfactorily explain their observed herbicidal activity and selectivity toward grasses, broadleaves, and crops. They concluded that a combination of steric, acidic, lipophilic, and electronic properties must govern the pre- and post-emergence herbicidal activity of the TFMS compounds.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Hansch Structure–Activity Correlations

Although the complex herbicidal activity and selectivity characteristics of the trifluoromethanesulfonanilides could not be correlated with single molecular properties of series members, we believed that an appro-

Table I. Pre-Emergence Herbicidal Activity of Substituted Trifluoromethane Sulfonanilides (5)[a]



	Four Grass Species,[b] Dosage in lb./acre			*Four Broadleaf Species,[c] Dosage in lb./acre*		
X	*10*	*5*	*2.5*	*10*	*5*	*2.5*
2-Cl	100	92	72	100	100	92
*3-Cl	0	0	0	100	100	70
*4-Cl	30	0	0	93	72	40
4-F	100	90	90	97	97	97
*3-NO_2	50	0	0	100	100	0
2-SCH_3	100	100	100	87	60	52
*4-SO_2CH_3	93	93	85	0	0	0
2-CH_3, 4-Cl	40	0	0	100	100	100
2-Cl, 4-CF_3	100	57	30	100	97	53
2,4-di-Cl	100	100	100	100	100	100
2,4-di-F	100	100	100	100	100	100

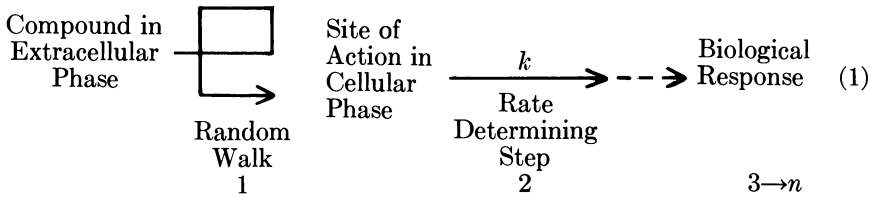
[a] Pre-emergence activity: 0 = no kill; 100 = complete kill. Tests run in illuminated greenhouse.

[b] Grass species: Giant Foxtail (*Setaria faberii*), Barnyard grass (*Echinochloa crus-galli*), Crabgrass (*Digitaria ischaemum*), Quackgrass (*Agropyron repens*).

[c] Broadleaf species: Pigweed (*Amaranthus retroflexus*), Purslane (*Portulaca vieracea*), Wild Mustard (*Brassica kaber*), Annual Morning Glory (*Ipomoea purpurea*).

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

priate combination of several molecular properties might explain the observed biological activity. In this connection, Hansch (*6, 7, 8, 9*) and Hansch and Fujita (*10, 11*) have developed a mathematical model for correlating biological activity with the physical parameters of molecules. Biological activity is correlated with the octanol/water partition coefficients and Hammett sigma (σ) constants (*12, 13*) characterizing the compounds being evaluated. The Hansch-Fujita model (*10*), which is usually applied to a series of biologically active compounds consisting of a parent molecule and its substituted derivatives, is illustrated in Equation 1.



(1)

Step 1 is a random walk process in which the biologically active molecule under investigation makes its way from a dilute solution outside a plant or animal cell to the site of action within the cell. The overall permeation event represented by Step 1 will presumably be a relatively slow process involving many partitions between lipid and aqueous phases in addition to adsorption and desorption onto tissues or proteins. Further, in most cases the partitioning process would probably depend considerably on the molecular structure and physical properties of the biologically active material. Once the compound has reached the receptor site or site of action within (or on the surface of) the cell, it is assumed that one rate-controlling step (Step 2) governs its biological activity and that Steps $3 \rightarrow n$ can be disregarded in expressing the rate of this reaction.

Hansch and Fujita used a Gaussian probability function to characterize the partitioning Step 1 and the Hammett function ($\log(k/k_o) = \rho\sigma$) to describe the rate Step 2 in their model (*10, 12, 13*). By appropriate mathematical treatment, they arrived at the following general structure–activity relationship which has come to be termed the Hansch Equation.

$$\log (1/C) = A\pi^2 + B\pi + \rho\sigma + D \tag{2}$$

In Equation 2, π is a comparative substituent constant which Hansch and Fujita (*10*) define as:

$$\pi = \log P_X - \log P_H \tag{3}$$

In Equation 3, P_H is the partition coefficient of an unsubstituted parent compound, and P_X is the corresponding partition coefficient of a substi-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

tuted derivative. As previously noted, the partition coefficients P are determined between 1-octanol and water in the Hansch correlation method. $\pi = (\log P_X - \log P_H)$ is a free-energy related parameter similar to the Hammett substituent constant σ (*12, 13*). Equation 2 shows that any family of biologically active compounds which exhibits a second-order dependence on π will also exhibit an optimum π value (π_o) which is characteristic of that particular series.

Biological activity is often reported in terms of a constant equivalent response (LD_{50}, LD_{90}, ED_{50}, percent growth, percent kill, etc.) obtained in a fixed time interval. Generally, this activity is expressed as the concentration (*e.g.*, LD_{50}, LD_{90}) necessary to cause a particular response in the given time interval. In an application of the Hansch equation aimed at correlating herbicidal activity data, C in Equation 2 would typically be an LD_{50} or LD_{90} value—*i.e.*, the molar extracellular concentration (application dosage) of the herbicide in question required to kill 50% or 90% of a particular weed type in a given time interval. C, π, and σ values are correspondingly read into a computer, and the parameters A, B, ρ, and D are calculated by stepwise regression analysis. In principle, once these fitting parameters have been determined for a given series of herbicidal, pesticidal, or pharmaceutical compounds in their action on a particular plant or animal system, the resulting equation can be used to predict the biological activity of other substituted compounds in the series from a knowledge of their octanol/water partition coefficients (π values) and Hammett substituent constants (σ values). Although unequivocal interpretation of final correlation equations determined *via* Hansch analysis is often difficult, clues to elucidate the possible mode of action of a particular active compound or related series of compounds are often available from the relative importance of the partitioning (π^2, π) and electronic (σ) terms in Equation 2. The Hammett σ constants are, of course, a measure of the electron-donating and -withdrawing power of aromatic ring substituents.

Objectives of the Present Study

The complex activity and selectivity characteristics of substituted trifluoromethanesulfonanilide herbicides were not correlatable with single molecular properties (*see* Table I and Ref. 5). We were particularly interested in determining whether the multiparameter Hansch equation, whose utility and apparent generality has already been substantiated in correlation studies on enzymatic systems (*14, 15, 16, 17, 25*), drugs and pharmaceuticals (*18, 19, 20, 23, 24*), pesticides (*10, 21*), and plant growth regulators (*22*), could be used to provide meaningful structure–activity

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

relationships among members of the TFMS series. The signs and magnitudes of the parameters A, B, ρ, and D would probably depend upon:

(a) the general hydrophobicity or hydrophilicity of the biologically active series of compounds being evaluated

(b) whether the series was being examined for herbicidal, pesticidal, or pharmaceutical activity

(c) whether formulation surfactants or emulsifiers were present which affected the partitioning characteristics of the series members

(d) the individual species of plant, animal, or insect on which the active compounds were being evaluated.

Item c was of particular interest since many surfactants are currently available for preparing herbicidal formulations for greenhouse and/or outdoor field testing.

The objectives of the current study were twofold and aimed at providing answers to the following questions. Can the Hansch biological correlations technique be used:

(1) to account for individual and/or group differences in the preemergence herbicidal activity and specificity of substituted trifluoromethanesulfonanilides toward grasses and broadleaf weeds?

(2) to assess quantitatively the effects of low levels of a common formulation surfactant (Tween 80) on the herbicidal activity of trifluoromethanesulfonanilides toward grasses and broadleaf weeds in terms of:

(a) changes in biological activity of individual series members?

(b) shifts in the range of optimum herbicidal activity from more lipophilic to more hydrophilic series members (or vice-versa)?

(c) overall changes in herbicidal potency of the trifluoromethanesulfonanilides (*e.g.*, general enhancement or inhibition of biological activity)?

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Experimental

Chemicals. The 1-octanol used in the partitioning experiments was high purity Eastman White Label reagent (#871, lot 691-1) and was optically transparent at all wavelengths $\geq$220 nm. Additional purification of the octanol *via* distillation was not necessary. Immediately before partitioning experiments, however, octanol samples were washed with NaOH and water as described below. Deionized, distilled water was used where necessary, and Mallinckrodt analytical reagent perchloric acid (#2766) was used to prepare the pH 1.0 aqueous phases for partitioning studies (*see below*).

All substituted trifluoromethanesulfonanilide (TFMS) samples were prepared in the 3M Co. Biochemical Research Laboratory and supplied by J. K. Harrington and R. D. Trepka of that laboratory. General preparative procedures for the TFMS compounds and related perfluoroalkanesulfonanilides have been outlined (*1, 2, 3, 4, 5*). Additional details of the sulfonylation procedures and subsequent modifications of the resulting products will be published later.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Tween 80 (also known as At-Plus 109 and Polysorbate 80) is a polyoxyethylene sorbitan monooleate and was obtained from Atlas Chemical Industries, Inc. This nonionic surfactant has an HLB (hydrophile-lipophile balance) of 15.0 and is used as an emulsifier, solubilizer, and dispersant. It was used without further purification at the 0.1% (w/v) level in partitioning and greenhouse herbicidal evaluations as described below.

Determination of Octanol/Water Partition Coefficients. Since octanol/water partition coefficient data for trifluoromethanesulfonanilide (I) and its substituted derivatives have not been reported in the literature and since it was not apparent that any of the substituent π values previously determined by Fujita *et al.* (*11*) would be directly applicable to the TFMS herbicidal system, all TFMS partition coefficients and π values were determined experimentally. The fluoroalkanesulfonanilides are very acidic because of the electron-withdrawing power of the parent fluoroalkanesulfonyl group (*5*). The parent TFMS compound (I), for example, has a pK_a in water of 4.45 at 25°C. This inherent acidity extends to all TFMS series members and requires that the usual partition coefficient measurement procedure described by Fujita *et al.* (*11*) be modified to obtain accurate values of log P and π.

The octanol/water partition coefficient P is defined as

$$P = \frac{C_o}{C_w\,(1-\alpha)} \tag{4}$$

where C_o is the concentration of the compound of interest in the octanol phase after partitioning, and C_w is its corresponding concentration in the water phase. The term α is the degree of dissociation of the acidic compound in the water phase and must be considered for all ionizable compounds. Since all the TFMS compounds in our study were ionizable in water, either the degree of dissociation (α) had to be determined for all members of the series under the pH conditions of the partitioning, or solution pH conditions had to be adjusted so that the compounds in question remain in the nonionized state ($\alpha \rightarrow 0$) in the aqueous phase during the partitioning experiment (*33*). The latter approach, in which we used a low, buffered pH, was chosen for this study. This technique was verified using the well-characterized benzoic acid system (*11*). At pH 7.0, for example, benzoic acid ($pK_a = 4.19$ at 25°C) is very soluble in water since it exists essentially entirely in its ionized state. Because of this ionization, octanol/water partitioning experiments at neutral pH are difficult to carry out and interpret. By conducting partitioning experiments at low pH (several units lower than the pK_a of benzoic acid—*e.g.*, pH 1.0), the acid is maintained in its nonionized state ($\alpha \rightarrow 0$), and to an excellent approximation the partition coefficient P is given by C_o/C_w (*cf.* Equation 4). Accordingly, octanol/water partitioning experiments were done on benzoic acid and two of its derivatives (*p*-chlorobenzoic acid and *m*-methoxybenzoic acid) using the procedure of Fujita *et al.* (*11*) except that their distilled-water phase was replaced with pH 1.0 water (0.1*N* perchloric acid) in our tests. As shown in Table II, the partition coefficients so determined at low pH for benzoic acid and the two

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

derivatives (using $P = C_o/C_w$) were in excellent agreement with the unbuffered aqueous (and necessarily higher pH) values reported by Fujita *et al.* (*11*) where the degree of dissociation α was used in the partition coefficient calculations.

Table II. Comparison of Partition Coefficients Using pH 1.0 *vs.* Distilled Water Aqueous Phases

Compound	*log* P; *pH 1.0 $HClO_4$ (this study)*	*log* P; *Distilled Water (Ref.* 11)
Benzoic Acid	1.86 ± 0.03	1.85 ± 0.03
p-Chlorobenzoic Acid	2.63 ± 0.03	2.72 ± 0.03
m-Methoxybenzoic Acid	2.00 ± 0.03	1.99 ± 0.03

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

All partition coefficient measurements on the 15 trifluoromethanesulfonanilides were carried out using a pH 1.0 water phase (acidified with perchloric acid). The pH of the acidic water phase remained essentially unchanged during partitioning. The pK_a values at 25°C in water of all but two of the TFMS derivatives (4-SO_2CH_3 and 3-SO_2CH_3 series members) were considerably greater than 3.0 so that all material partitioned into the pH 1.0 water phase was unionized. It was thus possible to use the simplified expression $P = C_o/C_w$ to calculate the partition coefficients. For the 4-SO_2CH_3– and 3-SO_2CH_3–TFMS derivatives, whose aqueous pK_a's are known to be close to 3.0, potentiometric titrations were carried out to establish accurately the pK_a values [$pK_{a,25°C}$ (4-SO_2CH_3–TFMS) = 2.80; $pK_{a,25°C}$ (3-SO_2CH_3–TFMS) = 3.10]. These values were then used to determine α at pH 1.0 with the partition coefficients calculated according to Equation 4.

The 1-octanol used in partitioning experiments was purified by washing three times with 1*N* NaOH, followed by six washings with distilled, deionized water. The initial high purity of the octanol eliminated a subsequent distillation step. For the partitioning studies, octanol saturated with pH 1.0 distilled water, and pH 1.0 distilled water (0.1*N* perchloric acid) saturated with octanol were used. As judged from our preliminary model studies at pH 1.0 on the benzoic acid series (*see above*), the presence of 0.1*N* perchloric acid in the aqueous phase had no significant effect on the ultimate partitioning behavior of either the benzoic acid or TFMS compounds examined.

In a typical partitioning experiment, 0.12 to 0.15 gram of the TFMS compound was dissolved in 50 ml of 1-octanol to a final concentration of from $5 \times 10^{-3}M$ to $1.3 \times 10^{-2}M$. The octanol solution of TFMS compound was partitioned against an equal volume of pH 1.0 water by agitating for at least an hour on a mechanical shaker. The mixture was then allowed to stand until the octanol and water layers separated. The aqueous layer was removed, centrifuged to remove any cloudiness, and analyzed by ultraviolet (UV) absorption spectrophotometry to determine the concentration of TFMS compound partitioned into the aqueous phase. The concentration of the TFMS compound remaining in the octanol phase was calculated by difference.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

All UV determinations were made on a Cary Model 14 spectrophotometer. The absorption maxima of the 15 partitioned TFMS compounds in water ranged from 259 nm for the 4-SCH_3 derivative to 214 nm for the 4-F and 3-Cl derivatives. All UV measurements were made in 1-cm cuvettes with typical partitioned TFMS compounds exhibiting optical densities ranging from 0.5 to 1.5 in the aqueous phase. Log *P* values determined by the above procedure for the TFMS compounds were precise to better than ±2%.

Partitioning in the Presence of Surfactant. Partitioning experiments in the presence of Tween 80 were carried out in a manner identical to that discussed above except that 0.1% (w/v) of the surfactant was thoroughly mixed with the water phase before partitioning. Octanol/water partitioning studies carried out with only 0.1% Tween 80 (and no TFMS compound) present initially indicated that at equilibrium, the surfactant partitions approximately 60/40 in favor of the octanol phase.

In a typical octanol/water partitioning experiment carried out in the presence of Tween 80, the TFMS compound was first dissolved in 50 ml octanol and then partitioned against an equal volume of pH 1.0 water containing 0.1% Tween 80 as described above. All samples were centrifuged to ensure the best possible octanol/water phase separation and optical clarity.

Equilibrium concentrations of TFMS compounds in the aqueous phase were again determined by UV. The concentration in the octanol phase was determined by difference, as usual. Although Tween 80 exhibited an absorption maximum at 232 nm, the optical density of the surfactant at this wavelength in the aqueous phase of the final partitioned samples usually did not exceed 0.2. Most of the TFMS compounds exhibited absorption maxima at wavelengths either near 215 ± 5 nm or near 255 ± 5 nm where the optical density of the partitioned Tween 80 was always $\leqslant$0.1. Absorption interference by the surfactant in wavelength regions used to measure the concentrations of the final partitioned TFMS compounds was therefore not a problem because of the low surfactant absorption in spectral regions of TFMS absorption. Typical TFMS optical densities at absorption maxima, for example, were 0.8 to 1.5. As a precaution, however, the aqueous phase of a control sample consisting of 0.1% Tween 80 partitioned as described above between octanol and pH 1.0 water was used as a reference (blank) in all UV determinations. All UV absorption at the wavelengths of interest could thus be ascribed to TFMS absorption within experimental error. It did not appear that surfactant-enhanced octanol partitioning into the water phase led to any significant errors in TFMS concentration determinations. For the TFMS compounds which partitioned much more strongly toward the water layer in the presence of Tween 80 than in its absence (*see* Table VI), an opalescent or cloudy aqueous phase was formed because of the emulsifying effect of the surfactant on the partitioned compound. In these cases, an aliquot of the final partitioned water phase was diluted 1:5 with pH 1.0 water; this cleared the solutions optically and permitted UV determination of the partitioned TFMS compound concentration. The precision of the log *P* values determined for the TFMS compounds in the presence of 0.1% Tween 80 is ±5%.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Greenhouse Tests. Pre-emergence herbicidal evaluations of the 15 TFMS compounds listed in Table IV were conducted in an artificially illuminated greenhouse. Two grass species (Giant Foxtail, *Setaria sp.;* Cheat Grass, *Bromus secalinus*) and a broadleaf weed (Wild Mustard, *Brassica kaber*) were used in the tests. For each TFMS evaluation, the three weed species were planted as seeds in separate rows in the same 6 inch diameter round plastic pot. The soil was a sterilized sandy loam containing 68% sand, 20% silt, 9% clay, and 2–3% organic matter. The soil was pasteurized at 180°F for 30 minutes before use, a procedure which killed pathogens but not beneficial soil bacteria.

Two series of replicated pre-emergence herbicidal tests were conducted simultaneously on the TFMS compounds. In the first, no surfactant was used in the herbicidal formulations; in the second, Tween 80 at the 0.1% (w/v) level was added to the formulations. All procedures and preparations were identical for both series except for the presence of the surfactant in formulations of the second series.

Standard herbicidal formulations in the absence of surfactant were prepared by dissolving the highest intended dose of each TFMS compound in a 1% acetone–water solution. The small amount of acetone (necessary to effect solution of some of the more lipophilic series members) was used in all formulations for consistency. Standard herbicidal formulations in the presence of surfactant were prepared in a similar fashion, except that the 1% acetone–water contained in addition 0.1% (w/v) Tween 80.

The pre-emergence herbicidal tests were done at three or four dosage levels ranging from 1.25 to 20 lb/acre for each TFMS derivative. The lb/acre designation was obtained by dividing the actual dry weight in grams of TFMS herbicide applied to each pot from an aqueous drench by the topsoil surface area in the pot ($\sim$0.2 ft^2) and then applying an appropriate conversion factor $\left(e.g., \frac{0.01 \text{ gram TFMS compound}}{0.2 \text{ ft}^2 \text{ per pot}} = \frac{4.9 \text{ lb}}{\text{acre}}\right)$. For subsequent computer correlation studies, the lb/acre data were converted to moles/acre. Herbicidal formulations corresponding to different dose levels were prepared by diluting aliquots of the standard herbicidal formulations with either 1% acetone–water (tests with no surfactant added) or 1% acetone–water containing 0.1% Tween 80 (tests with surfactant added). Thus, in the surfactant tests, Tween 80 was kept at a constant concentration of 0.1% at all herbicidal dosages. A total of 80 ml of each herbicidal formulation were applied to each pot as a single drench immediately after seeds were planted in the pre-emergence evaluations. Thereafter, pots were bottom-watered daily until plants emerged. Top watering was then used for the rest of the test.

Pre-emergence herbicidal test data were evaluated after 8, 21, and 42 days for both test series. The herbicidal response of each TFMS compound at each time interval was rated (in terms of % kill by comparing with a control sample containing no herbicidal treatment) on a 0–100 scale (where 0 = no activity, 50 = 50% kill, 100 = complete kill, etc.). No significant changes in relative herbicidal activity were observed after 21 days. The 21-day herbicidal activity ratings were used to prepare log-probit plots from which LD_{90} values were determined for each TFMS compound in the presence and absence of Tween 80 (*see* Results).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

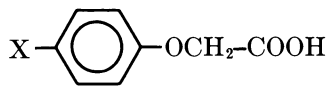
In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

All herbicidal tests were carried out simultaneously. Errors and variations from such uncontrollable factors as seasonal changes in plant growth were thus minimized.

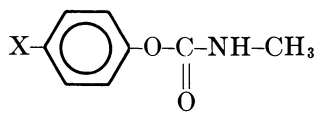
Computer Correlations. All correlations of herbicidal test data with π- and σ-parameters according to Equation 2 were carried out with an IBM 360 computer using a least-squares stepwise regression program of considerable versatility (UCLA Health Sciences Program BMDO2R). Multiple correlation coefficients (r), parameter and equational standard errors, and F tests were routinely used to judge the goodness of fit of herbicidal activity data to the Hansch equation. Three-dimensional perspective plots of final equational forms were carried out on a Calcomp plotter using a program written by Earl Cook (3M Co.).

Results

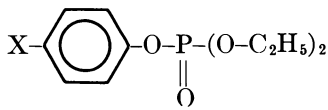
Model Parent Compound Series. Experimental partition coefficient data for a variety of substituted benzenes and seven other related parent compound series (phenoxyacetic acid, phenylacetic acid, benzoic acid, benzyl alcohol, phenol, aniline, nitrobenzene) were reported in 1964 by Fujita *et al.* (*11*). The π values (*see* Equation 3) derived for individual substituents in each of the above-mentioned parent compound series have since been frequently used (with varying degrees of success) by many investigators to approximate π values for the corresponding substituents in other related parent compounds for which no experimental partitioning data are available. For example, Hansch and Deutsch (*26*), in a correlation study of structure–activity relationships in cholinesterase inhibitors, used π values derived for aromatic ring substituents (X) in the phenoxyacetic acid series



to approximate π values for the same substituents in parent series of methylcarbamates:

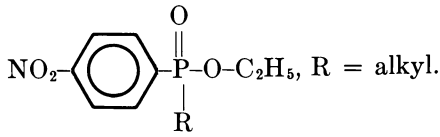


diethylphenylphosphates:

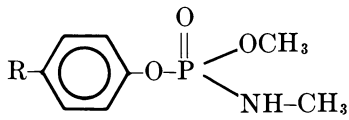


Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

and even alkylphosphonic acid esters:



Neely and Whitney (*21*) have similarly used the phenoxyacetic acid series π values in structure–activity correlations of the insecticidal activity of phenyl-*O*-methyl methylphosphoramidates:



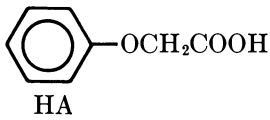
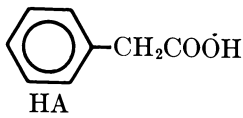
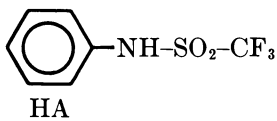
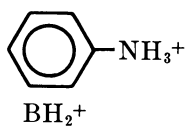
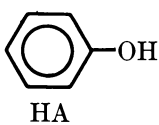
Before experimentally measuring the partition coefficients of our 15 trifluoromethanesulfonanilide (TFMS) series members, we too had given serious consideration to approximating the TFMS substituent π values with appropriately chosen π values from one of the above-mentioned parent series studied by Fujita *et al.* (*11*). The appropriate choice of an approximating parent compound was difficult to make, however, as can be judged from the ionization and partitioning data tabulated in Table III for several reasonable model series chosen from Ref. *11*.

The five parent compounds in Table III are arranged in order of increasing pK_a of their ionizable protonic groups. For phenoxyacetic acid ($pK_a = 3.17$) and phenylacetic acid ($pK_a = 4.31$), the primary ionization is that of the carboxylic acid side chain. The acidity of the TFMS parent compound ($pK_a = 4.45$) is attributable to the loss of the relatively labile proton from the parent side chain (ϕ–NH–SO_2–$CF_3 \rightleftarrows \phi$–$N^-$–$SO_2$–$CF_3 + H^+$). For aniline, the process with $pK_a = 4.63$ is associated with protonic ionization of the anilinium cation. The $pK_a = 9.89$ process in phenol refers to the formation of phenolate anion.

Although the pK_a's of phenoxyacetic acid, phenylacetic acid, TFMS, and aniline are all quite similar, differing by less than 1.5 p*K* units for the most extreme comparison (phenoxyacetic acid *vs.* aniline), this similarity ends when the partitioning behavior of these same parent compounds is compared. Whereas the logarithm of the octanol/water partition coefficient (P) varies only from log $P = 0.90$ for aniline to log $P = 1.41$ for phenylacetic acid, TFMS (log $P = 3.05$) is 1.6 orders of magnitude more lipophilic than phenylacetic acid, the most hydrophobic of the other parent compounds exhibiting a similar pK_a. If the π values for the side chains of the parent compounds listed in Table III are calculated using Equation 5 as shown on p. 196,

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Table III. Choice of Model Parent Series

Parent Molecule	*Ionization Step*	*Partitioning*[a,b]
1. Phenoxyacetic Acid (POA)  HA	$HA \rightleftharpoons H^+ + A^-$ $pK_a = 3.17$ (25°C)	$\log P = 1.27$ $\pi(OCH_2COOH) = -0.86$
2. Phenylacetic Acid (PAA)  HA	$HA \rightleftharpoons H^+ + A^-$ $pK_a = 4.31$ (25°C)	$\log P = 1.41$ $\pi(CH_2COOH) = -0.72$
3. Trifluoromethane-sulfonanilide (TFMS)  HA	$HA \rightleftharpoons H^+ + A^-$ $pK_a = 4.45$ (25°C)	$\log P = 3.05$ $\pi(NHSO_2CF_3) = +0.92$
4. Aniline  BH_2^+	(a) $BH_2^+ \rightleftharpoons BH + H^+$ $pK_a = 4.63$ (25°C)	$\log P = 0.90$ $\pi(NH_2) = -1.23$
	(b) $BH \rightleftharpoons B^- + H^+$ $pK_a \approx 27$ (25°C)	
5. Phenol  HA	$HA \rightleftharpoons H^+ + A^-$ $pK_a = 9.89$ (25°C)	$\log P = 1.46$ $\pi(OH) = -0.67$

[a] Partitioning data for parent compounds 1, 2, 4, and 5 obtained from Ref. *11*.
[b] Partitioning data for 3 measured in this study.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

$$\pi \text{ (parent side chain)} = \log P \text{ (parent compound)} - \log P \text{ (benzene)}$$
$$= \log P \text{ (parent compound)} - 2.13 \quad (5)$$

the hydrophobicity of the $-NHSO_2CF_3$ side chain in TFMS is illustrated even more graphically. The large negative π values characterizing the side chains of phenoxyacetic acid, phenylacetic acid, aniline, and phenol indicate that their substitution into the parent benzene ring renders the latter significantly more hydrophilic and water soluble (over an order of magnitude for $-NH_2$ substitution). The substitution of $-NHSO_2CF_3$ into the benzene ring contrastingly renders the latter almost an order of magnitude more lipophilic and soluble in octanol. The gross difference in lipophilicity of the trifluoromethanesulfonanilide side chain relative to the much more hydrophilic side chains of the other parent compounds strongly suggests that similarities in the side chain pK_a's (and hence overall parent molecule acidities) should not be used as a basis for selecting the substituent π values of another parent series to approximate substituent π values in the TFMS series.

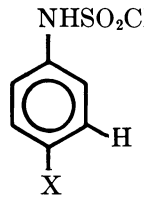
One might instead choose as a basis for comparison the relative electron-withdrawing or -donating properties of the parent side chains in Table III. Both the $-NH_2$ side chain of aniline and the $-OH$ group in phenol are strong donators of electrons to the benzene ring. The strong electron-withdrawing properties of the trifluoromethanesulfonyl portion of the TFMS side chain, however, would be expected to mitigate drastically the corresponding electron-donating tendencies of the nitrogen attached to the benzene ring in TFMS. In addition to the differences in partitioning properties, it would also appear unwise to choose either a phenol or aniline parent series to represent the TFMS series because of the expected differences in electron-donating properties of the parent side chains [which might well be expected to influence (to a greater or lesser degree) the hydrophilicity or hydrophobicity of other substituents that might be placed in the benzene ring]. Therefore, even though TFMS is a substituted aniline of sorts, there is little basis for comparing it with aniline. The amino hydrogen in aniline corresponding to the labile hydrogen ($pK_a = 4.45$) in TFMS, for example, has an extremely high pK_a of 27 (*see* second ionization equilibrium for aniline in Table III), further illustrating the dissimilarity of the two parent molecules. In addition, the combined electron-withdrawing tendencies of the phenoxy and carboxyl (in phenoxyacetic acid) and phenyl and carboxyl (in phenylacetic acid) do not render either of the methylene ($-CH_2-$) hydrogens in the side chains of these parent molecules particularly acidic. This is in contrast to the combined electron-withdrawing effect of phenyl and trifluoromethanesulfonyl on the –NH linkage in the side chain of TFMS which is highly acidic.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

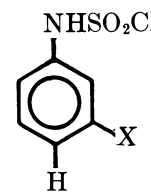
Therefore, it is difficult to define a reasonable basis for selecting any of the parent series previously examined by Fujita *et al.* (*11*) as a model for the TFMS series of the present study. Indeed, some initial Hansch analyses of our TFMS pre-emergence herbicidal activity data using Fujita's phenoxyacetic acid substituent π values produced very poor correlations. We thus deemed it prudent (if not essential) to determine experimentally π values for all substituents in the TFMS series.

Table IV. Trifluoromethanesulfonanilides Chosen for Herbicide Test[a]



4-Substituted

X	π[b]	σ
4-CF_3	+1.42	+0.551
4-Cl	+0.91	+0.227
4-SCH_3	+0.69	−0.047
4-CH_3	+0.55	−0.170
4-F	+0.20	+0.062
4-OCH_3	−0.05	−0.268
4-SO_2CH_3	−1.06	+0.728
H	0.00	0.000



3-Substituted

X	π[b]	σ
3-CF_3	+1.45	+0.415
3-Cl	+0.95	+0.373
3-F	+0.37	+0.337
3-SCH_3	−0.20	+0.144
3-$COCH_3$	−0.27	+0.306
3-OH	−0.54	−0.002
3-SO_2CH_3	−1.20	+0.647

[a] No Tween 80 present.
[b] *See* Table V for standard deviation in π values.

π Values in the Absence of Surfactant. Table IV gives π and σ values for the 15 trifluoromethanesulfonanilides examined in this study. Octanol/water partition coefficients in the absence of surfactant were determined as described, and substituent π values were calculated according to Equation 3 using log $P_H = 3.05$ for the H-substituted parent compound (*see* Table III). Hammett sigma constants (σ) were taken from the compilation of Jaffe (*13*). Our study was restricted to meta- and para-monosubstituted TFMS derivatives. Ortho-substituted TFMS derivatives were not included in the herbicidal tests to eliminate the usual difficulties associated with selection and interpretation of Hammett σ constants for 2-substituents. The substituted TFMS derivatives in Table IV were chosen to span a broad spectrum of π and σ values. π values ranged from +1.42 to −1.06 for the 4-substituted TFMS derivatives and from +1.45 to −1.20 for the 3-substituents. Corresponding σ constants likewise covered a fairly broad range of values for both 3- and 4-substituents. For

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

both the meta- and para-substituted TFMS series, the most hydrophobic substituent was trifluoromethyl ($-CF_3$) while methylsulfonyl ($-SO_2CH_3$) was the most hydrophilic side chain in each series.

Table V gives substituent π values measured in the absence of surfactant for the five parent series previously compared in Table III. Examination of the table indicates that the experimentally measured TFMS series π values do not coincide with those of any of the other parent series but lie roughly midway between the corresponding π values of the phenoxyacetic acid and phenol series. In those cases where comparisons are possible, π values for the 4-CH_3, 4-F, 4-OCH_3, 3-$COCH_3$, and 3-OH substituents in the TFMS series lie closer to the π values of the corresponding substituents in the phenoxyacetic acid series than they do to those in the phenol series. In contrast, π values for TFMS 4-Cl, 3-CF_3, 3-Cl, and 3-F substituents lie closer to corresponding phenol series π values. On the basis of the direct comparison between parent series afforded by Table V, it is evident that none of the partitioning data previously presented by Fujita *et al.* (*11*) for the various model parent series adequately describe the partitioning behavior of the substituted trifluoromethanesulfonanilide herbicides.

Effect of Tween 80 on TFMS π Values. Table VI lists octanol/water partitioning data for the 15 substituted TFMS compounds used in this study. Logarithms of the partition coefficients obtained both in the absence ($\log P_X$) and presence ($\log P_X'$) of 0.1% (w/v) Tween 80 are given. (Procedures for determining the partition coefficients P_X and P_X' have been outlined in the Experimental section.) π, π', and π'' values in Table VI were calculated according to the following relationships:

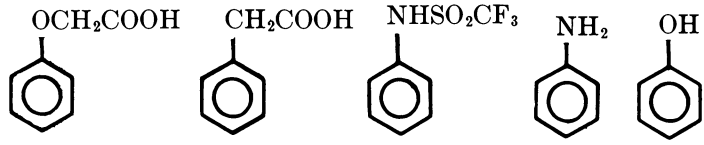
$$\begin{aligned} \pi &= \log P_X - \log P_H && \text{(no Tween 80 present)} \\ \pi' &= \log P_X' - \log P_H && \text{(0.1\% Tween 80 present)} \\ \pi'' &= \log P_X' - \log P_H' && \text{(0.1\% Tween 80 present)} \end{aligned} \qquad (6)$$

In Equation 6, π and π' values are calculated relative to $\log P_H = 3.05$ for the parent compound (*H*-substituted TFMS) determined in the absence of surfactant. Thus, the π' value determined for the parent compound in the presence of 0.1% Tween 80 differs from zero ($\pi'_H = -0.45$). The π'' substituent constants in Table VI and Equation 6 are calculated relative to the logarithm of the partition coefficient of the parent TFMS compound determined in the presence of 0.1% Tween 80 ($\log P_H' = 2.60$).

Since Tween 80 itself partitions 60/40 in favor of the octanol phase, a significant effect of this surfactant on the partitioning properties of the TFMS series members might be expected. Table VI indicates this to be the case. Comparing $\log P_X$ with $\log P_X'$ values, it is evident that the

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Table V. Substituent π Values for Models[a,b]



Substituent (X)	*POA* pK_a = *3.17* π (X)	*PAA* pK_a = *4.31* π (X)	*TFMS* pK_a = *4.45* π (X)	*Aniline* pK_a = *4.63* π (X)	*Phenol* pK_a = *9.89* π (X)
1. H	0.00	0.00	0.00	0.00	0.00
2. 4-CF_3	—	—	+1.42 ±0.03	—	—
3. 4-Cl	+0.70 ±0.03	+0.70 ±0.04	+0.91 ±0.02	—	+0.93 ±0.01
4. 4-SCH_3	—	—	+0.69 ±0.02	—	—
5. 4-CH_3	+0.52 ±0.05	+0.45 ±0.03	+0.55 ±0.02	+0.49 ±0.02	+0.48 ±0.01
6. 4-F	+0.15 ±0.01	+0.14 ±0.01	+0.20 ±0.01	+0.25 ±0.02	+0.31 ±0.01
7. 4-OCH_3	−0.04 ±0.01	+0.01 ±0.02	−0.05 ±0.01	—	−0.12 ±0.01
8. 4-SO_2CH_3	—	—	−1.06 ±0.02	—	—
9. 3-CF_3	+1.07 ±0.02	+1.16 ±0.03	+1.45 ±0.03	—	+1.49 ±0.01
10. 3-Cl	+0.76 ±0.02	+0.68 ±0.03	+0.95 ±0.02	+0.98 ±0.02	+1.04 ±0.01
11. 3-F	+0.13 ±0.03	+0.19 ±0.02	+0.37 ±0.01	+0.40 ±0.02	+0.47 ±0.01
12. 3-SCH_3	+0.62 ±0.01	—	−0.20 ±0.02	—	—
13. 3-$COCH_3$	−0.28 ±0.01	—	−0.27 ±0.01	—	−0.07 ±0.01
14. 3-OH	−0.49 ±0.01	−0.52 ±0.02	−0.54 ±0.02	−0.73 ±0.04	−0.66 ±0.01
15. 3-SO_2CH_3	−1.26 ±0.05	−1.25 ±0.01	−1.20 ±0.03	—	—

[a] No Tween 80 present.
[b] Partitioning data for POA, PAA, aniline, and phenol taken from Ref. *11*.

latter quantities have been substantially reduced in the presence of Tween 80, indicating a decrease in hydrophobicity and increase in hydrophilicity of the original TFMS series members. Except for the 3-CF_3–TFMS derivative, the π' values in Table VI indicate that the remaining 14 series members (including the parent compound) are significantly more hydrophilic in the presence of Tween 80 than was the original parent compound

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

in the absence of surfactant. Contrastingly, before the addition of Tween 80, only six of the original 15 TFMS derivatives (*i.e.*, 4-OCH_3, 4-SO_2CH_3, 3-SCH_3, 3-$COCH_3$, 3-OH, and 3-SO_2CH_3–TFMS) were less lipophilic than the parent species.

Table VI. Effect of 0.1% Tween 80 on Octanol/Water Partitioning Characteristics of TFMS Herbicides[d]

	No Surfactant		*0.1% Tween 80*		
Substituent	*Log* P_X	π	*Log* P_X'	π'[b]	π''[c]
1. H	3.05	0.00	2.60	−0.45	0.00
2. 4-CF_3	4.47	+1.42	2.90	−0.15	+0.30
3. 4-Cl	3.96	+0.91	2.60	−0.45	0.00
4. 4-SCH_3	3.74	+0.69	2.70	−0.35	+0.10
5. 4-CH_3	3.60	+0.55	2.50	−0.55	−0.10
6. 4-F	3.25	+0.20	2.50	−0.55	−0.10
7. 4-OCH_3	3.00	−0.05	2.30	−0.75	−0.30
8. 4-SO_2CH_3	1.99	−1.06	1.80	−1.25	−0.80
9. 3-CF_3	4.50	+1.45	3.20	+0.15	+0.60
10. 3-Cl	4.00	+0.95	[a]	[a]	[a]
11. 3-F	3.42	+0.37	2.80	−0.25	+0.20
12. 3-SCH_3	2.85	−0.20	2.50	−0.55	−0.10
13. 3-$COCH_3$	2.78	−0.27	2.50	−0.55	−0.10
14. 3-OH	2.51	−0.54	2.10	−0.95	−0.50
15. 3-SO_2CH_3	1.85	−1.20	1.70	−1.35	−0.90

[a] Water phase was too cloudy to measure partition coefficient spectrophotometrically.
[b] $\pi' = \log P_{X'} - \log P_H$.
[c] $\pi'' = \log P_{X'} - \log P_{H'}$.
[d] *See* text for definitions and details.

Referring back to Table VI, note that the partition coefficients of those TFMS series members substituted with relatively hydrophobic groups (*e.g.*, 4-CF_3, 3-CF_3) are significantly more affected by the presence of Tween 80 than are hydrophilically substituted sulfonanilides (*e.g.*, 4-SO_2CH_3, 3-SO_2CH_3, 3-OH). The surfactant thus acts primarily to water-solubilize the most hydrophobic of the TFMS series members. The net effect of this increased hydrophobic solubility in the presence of Tween 80 is to compress the range of substituent π' values relative to their original spread in the absence of surfactant. All TFMS derivatives correspondingly become more uniformly soluble in water when the surfactant is added to the partitioning mixture.

The π'' values in the far right column of Table VI bring to light another interesting fact. They indicate that Tween 80, in addition to lowering the octanol/water partition coefficients of all TFMS series members in a general fashion, appears to alter the order of lipophilicity of certain individual series members. In the presence of surfactant, for example, the 4-Cl, 4-CH_3, and 4-F TFMS derivatives all exhibit equal or

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

enhanced hydrophilicity relative to the H-substituted parent compound. In the absence of surfactant, each of these derivatives is considerably more lipophilic than the parent compound. Hence, one must conclude that the water-solubilizing action of Tween 80 on the 15 TFMS compounds involves more than just a uniform shift in the hydrophobicity of the various series members. Rather, the surfactant may interact in a unique fashion with at least several of the member side chains to form unusually water-soluble micellar species.

Evaluation of TFMS Herbicidal Activity. The herbicidal potency of the 15 substituted TFMS compounds was rated after a 21-day test period on a 0–100% kill scale. Herbicidal test data were collected for two grass species (Foxtail, Cheat Grass) and a broadleaf weed (Wild Mustard) in the presence and absence of 0.1% (w/v) Tween 80. Since the tests

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

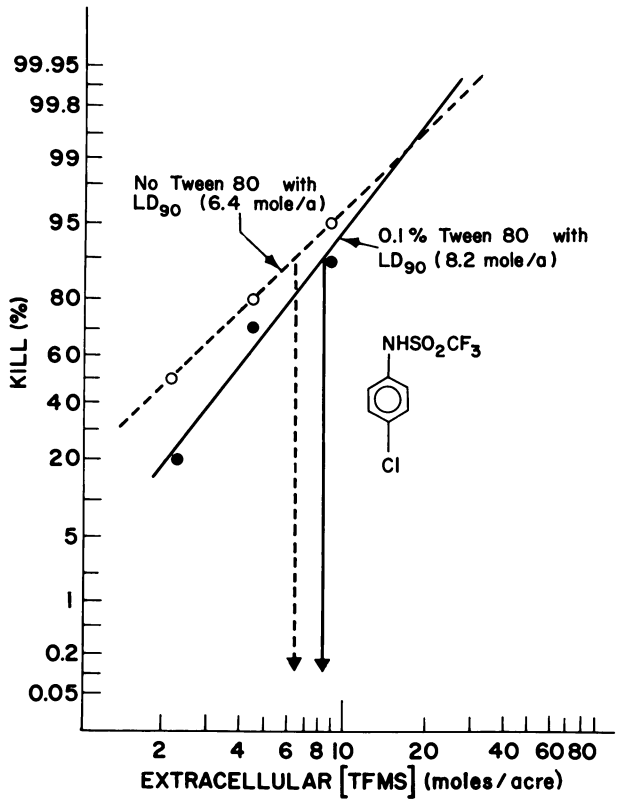


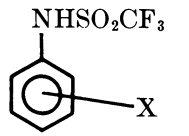
Figure 1. Log-probit plot for the 4-Cl-TFMS derivative acting on Foxtail Grass

○ *Data points obtained in the absence of 0.1% Tween 80*
● *Data points obtained in the presence of 0.1% Tween 80*

were carried out at three or four dosage levels for each TFMS compound, it was possible to construct log-probit plots from which LD_{90} values characterizing the potency of each herbicide could be determined. The theory and method of application of log-probit analysis has been presented elsewhere (*27, 28, 29, 30*). In Figure 1 is presented a typical log-probit plot for Foxtail grass illustrating the method of determining LD_{90} concentrations for the 4-Cl TFMS derivative in the presence and absence of 0.1% Tween 80. The LD_{90} value for each of the test compounds was defined as the molar extracellular concentration (in moles/acre) required to kill 90% of the weed type under consideration in pre-emergence tests carried out over a 21-day time interval. In all cases, a calculated least-squares straight line was drawn through the experimental points on the log-probit plot. Individual data points on the plots were the average % kill values of the two replicated experiments conducted at each herbicidal dosage. Replicates generally agreed in activity to better than ±5%. Since data points typically fell within the 10–90% kill range, no statistical weighting factors were applied to the points in constructing the plots. The LD_{90} value for each herbicide was taken as the extracellular TFMS concentration (abscissa) corresponding to the 90% kill point (ordinate) on the fitted log-probit line.

Table VII lists the LD_{90} values determined for each of the 15 TFMS derivatives in pre-emergence tests on Foxtail, Cheat Grass, and Wild Mustard. The Foxtail and Wild Mustard tests were carried out in the

Table VII. LD_{90} Values for TFMS Pre-Emergence Herbicides



Substituent (X)	*Cheat Grass* LD_{90}, *mole/acre No Tween 80*	*Foxtail Grass* LD_{90}, *mole/acre No Tween 80*	*Foxtail Grass* LD_{90}', *mole/acre 0.1% Tween 80*	*Wild Mustard (Broadleaf)* LD_{90}, *mole/acre No Tween 80*	*Wild Mustard (Broadleaf)* LD_{90}', *mole/acre 0.1% Tween 80*
1. H	25.0	8.0	12.7	4.6	6.4
2. 4-CF_3	4.05	9.7	10.6	3.35	2.64
3. 4-Cl	1.95	6.4	8.15	1.36	1.35
4. 4-SCH_3	2.58	5.1	5.1	42.0	25.2
5. 4-CH_3	43.7	82.0	40.0	18.9	10.5
6. 4-F	3.7	8.0	10.1	1.85	2.31
7. 4-OCH_3	80.0	58.5	23.0	35.6	80.5
8. 4-SO_2CH_3	2.21	2.42	2.64	31.6	79.0
9. 3-CF_3	22.6	15.0	21.5	6.15	1.95
10. 3-Cl	24.2	7.8	14.4	5.2	5.2
11. 3-F	23.0	8.35	18.2	5.55	4.67
12. 3-SCH_3	<2.1	<2.1	2.06	9.5	17.2
13. 3-$COCH_3$	29.7	14.6	12.4	24.3	33.0
14. 3-OH	107.0	128.0	204.0	46.2	32.9
15. 3-SO_2CH_3	<1.9	1.5	1.17	17.4	26.8

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

presence and absence of surfactant while Cheat Grass studies were conducted only in the absence of Tween 80. Examining first the herbicidal activity data gathered in the absence of surfactant, we note (Table VII) that significant differences in grass and broadleaf selectivity and overall activity are exhibited by many of the TFMS compounds. Note also the differences in activity of the individual series members toward the two grasses, Foxtail and Cheat Grass. The effect of Tween 80 on the Foxtail and Wild Mustard activity of the TFMS derivatives is particularly interesting. For Foxtail, the activity of nine of 15 of the sulfonanilides is inhibited by the presence of surfactant (higher LD_{90} values), the activity of four derivatives is enhanced, while two series members are relatively unaffected by Tween 80. For Wild Mustard, surfactant inhibits the activity of seven of the TFMS compounds, enhances the activity of six others, and has no effect on the activity of the 3-Cl– and 4-Cl–TFMS derivatives. For many of the TFMS derivatives listed in Table VII, Tween 80 exerts exactly opposite effects on the grass Foxtail and the broadleaf Wild Mustard.

Hansch Correlations of TFMS Herbicidal Activity. Stepwise regression techniques were used to correlate the pre-emergence herbicidal activity data gathered for all three weed types with one or more of the following appropriate general forms of the Hansch equation (*cf.* Equations 2 and 6 and Table VI).

$$\log (1/LD_{90}) = A\pi^2 + B\pi + \rho\sigma + D \quad (7)$$

$$\log (1/LD_{90}') = A\pi'^2 + B\pi' + \rho\sigma + D \quad (8)$$

$$\log (1/LD_{90}'') = A\pi''^2 + B\pi'' + \rho\sigma + D \quad (9)$$

π^2, π, and σ terms were added stepwise during the fitting procedure, the order of inclusion of these terms being determined on the basis of sequential statistical F tests.

For all regression analyses, herbicidal activity data were taken from the appropriate column in Table VII. Corresponding π, π' or π'' values were selected from Table VI. Hammett sigma constants (σ) were taken from the compilation of Jaffe (*13*) and correspond to those in Table IV. Since it was assumed throughout that σ would be relatively unaffected by the presence of surfactant, the σ values in Table IV were used to correlate data obtained in the presence and absence of Tween 80 for all three weed types. This assumption is reasonable since the surfactant was used at a low 0.1% level in all herbicidal and partitioning tests. Furthermore, surfactant effects would be expected to manifest themselves primarily in the partitioning behavior (π values) of the TFMS compounds

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

rather than in the Hammett σ constants (which are essentially functions of the electronic properties of the TFMS substituents).

Separation of TFMS Herbicides into Meta- and Para-Substituted Series. In our initial correlations, herbicidal activity data from Table VII for each weed type in the presence and absence of Tween 80 were fitted to the appropriate form of Equations 7–9. A typical result is illustrated by the stepwise regression obtained for Foxtail grass in the absence of surfactant:

Foxtail—No Tween 80 (n = *13*)

	r	r^2	SE	F	
$\log(1/LD_{90}) = 1.423\sigma - 1.436$ (± 0.340)	0.784	0.614	± 0.363	17.51	(10)
$\log(1/LD_{90}) = -0.249\pi^2 + 1.861\sigma - 1.366$ (± 0.189) (± 0.468)	0.819	0.671	± 0.351	10.21	(11)
$\log(1/LD_{90}) = -0.265\pi^2 + 0.017\pi + 1.893\sigma - 1.366$ (± 0.248) (± 0.157) (± 0.577)	0.820	0.672	± 0.370	6.14	(12)

In Equations 10–12, r is the multiple correlation coefficient, r^2 gives the percent correlation (*i.e.*, the fraction of the experimental data accounted for by the given regression equation), SE is the standard error of the equation (*i.e.*, the error in the fitted $\log(1/LD_{90})$ values, and F is the ratio of the mean sum of error squares removed by regression to the mean sum of squares of the error residuals not removed by regression. The F values are used in statistical tests to determine the goodness of fit of the given equation. The numbers in parentheses beneath the fit parameters in each equation denote the standard error in the respective parameters; they can also be used to judge the goodness of fit of a particular equation to the experimental data.

For reasons discussed below, data points for the 4-SCH_3 and 3-SCH_3 derivatives were omitted from the fits of Equations 10–12. Hence n, the number of data points included in the fits, equals 13 instead of 15. Examination of Equations 10–12 indicates, however, that even though $\log(1/LD_{90})$ exhibits a fairly strong dependence on σ, the data are not particularly well correlated by the Hansch equation. Addition of π^2 and π terms to Equation 10 raises the percent correlation (r^2) from 61 to 67%, but F is decreased and the standard error SE increased in the process. Hence, the correlation is not statistically improved.

Some initial graphical plots of $\log(1/LD_{90})$ *vs.* π and σ for Foxtail and the other weed types showed a grouping of data points which sug-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

gested that the 3- and 4-substituted sulfonanilides might in fact be functioning as two separate families of herbicides. Accordingly, herbicidal activity data for the TFMS compounds in Table VII were divided into two groups, those containing only 3-substituted and those containing only 4-substituted TFMS compounds. Each data class was then fitted separately to Equation 7. Since the proper classification of the H-substituted parent compound (I) was not immediately evident, it was omitted from the fits. The stepwise results obtained for Foxtail in the absence of surfactant for the two classes of 3- and 4-substituted TFMS herbicides are tabulated in Equations 13–18.

Foxtail—No Tween 80
Only 4-Substituted TFMS Derivatives (n = 6)

	r	r^2	SE	F	
$\log (1/LD_{90}) = 1.300\sigma - 1.372$ (± 0.377)	0.865	0.749	± 0.332	11.93	(13)
$\log (1/LD_{90}) = -0.490\pi^2 + 2.092\sigma - 1.169$ (± 0.264) (± 0.520)	0.940	0.883	± 0.262	11.32	(14)
$\log (1/LD_{90}) = -1.013\pi^2 + 0.332\pi + 3.032\sigma - 1.079$ (± 0.344) (± 0.179) (± 0.638)	0.978	0.957	± 0.194	14.81	(15)

Foxtail—No Tween 80
Only 3-Substituted TFMS Derivatives (n = 6)

	r	r^2	SE	F	
$\log (1/LD_{90}) = 2.878\sigma - 2.069$ (± 0.394)	0.965	0.930	± 0.394	53.47	(16)
$\log (1/LD_{90}) = -0.218\pi^2 + 3.373\sigma - 2.060$ (± 0.068) (± 0.267)	0.992	0.984	± 0.102	92.90	(17)
$\log (1/LD_{90}) = -0.223\pi^2 + 0.008\pi + 3.386\sigma - 2.062$ (± 0.094) (± 0.064) (± 0.344)	0.992	0.984	± 0.124	41.63	(18)

The statistical parameters in Equations 13–15 for the 4-substituted TFMS series members and Equations 16–18 for the 3-substituted compounds demonstrate that the fit of the biological data to the Hansch equation is significantly improved when the TFMS herbicides are separated into two classes (*cf.* Equations 10–12). Both the 3- and 4-series exhibit a strong dependence not only on σ, but also on π. For the 4-sub-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

stituted compounds both the π^2 and π terms are statistically significant in the Hansch equation while only the π^2 term is important for the 3-substituted TFMS series members (*cf.* Equations 17 and 18). In contrast to the correlations of Equations 10–12, where the importance of the partitioning parameter π was almost completely obscured when all data points were considered simultaneously, Equations 13–18 demonstrate that lipophilic partitioning and the relative hydrophobicity of TFMS series members are also of considerable importance in determining their overall herbicidal activity.

In a manner similar to that outlined above for Foxtail without surfactant, the TFMS compounds were separated into two groups consisting of only meta- or only para-substituted derivatives, and the corresponding data gathered in each herbicidal test series for the three weed types (in the presence and absence of Tween 80) were fitted to the Hansch equation. 4-SCH_3 and 3-SCH_3 data points were omitted from all the fits (*see below*). The H-substituted parent compound was omitted when the TFMS compounds were separated into two groups (*see above*). For comparison with Equations 10–18, the final best-fit correlation equations obtained for Foxtail in the presence of 0.1% Tween 80 are tabulated below.

Foxtail—0.1% Tween 80

All 3- and 4-Substituted TFMS Derivatives+Parent ($n=12$)

	r	r^2	SE	F	
	0.730	0.533	±0.447	3.05	(19)

$$\log (1/LD_{90}') = \underset{(\pm 0.924)}{-0.310\pi'^2} \underset{(\pm 1.190)}{- 0.720\pi'} \underset{(\pm 0.629)}{+ 1.278\sigma} - 1.690$$

Only 4-Substituted TFMS Derivatives ($n = 6$)

$$\log (1/LD_{90}') = \underset{(\pm 1.314)}{-2.443\pi'^2} \underset{(\pm 1.830)}{- 3.745\pi'} \underset{(\pm 0.482)}{+ 1.656\sigma} - 2.473$$

0.972	0.944	±0.152	11.17	(20)

Only 3-Substituted TFMS Derivatives ($n = 5$)

$$\log (1/LD_{90}') = \underset{(\pm 0.210)}{-0.776\pi'^2} \underset{(\pm 0.257)}{- 1.199\pi'} \underset{(\pm 0.242)}{+ 3.814\sigma} - 2.736$$

0.999	0.997	±0.081	127.76	(21)

Equations 19–21 again illustrate the importance of separating the TFMS series members into meta- and para-substituted groupings. The resulting improvement in statistical correlation is just as dramatic when surfactant is present in the herbicidal formulations as when it is absent.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

The conclusions reached above for Foxtail in the presence and absence of surfactant were not limited to this weed type. Similar correlations were carried out for Wild Mustard (in the presence and absence of surfactant) and for Cheat Grass (in the absence of Tween 80). In every case, correlations of π^2, π, and σ with $\log(1/LD_{90})$ were significantly improved when the TFMS herbicides were separated into two groups containing only meta- or para-substituted compounds and fitted individually. Table VIII lists the statistical parameters for the final best-fit equations which verify the above statements for each of the weed types and surfactant conditions evaluated. On the basis of the decreases in *SE* and the significant increases in *r*, r^2, and *F*, one must conclude from the data in Table VIII that separation of the TFMS herbicides into 3- and 4-substituted derivatives for correlating structure with activity *via* regression analysis is well justified. We have thus used this classification procedure throughout our study.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Table VIII. Effect on Statistics of Hansch Correlation of Separating TFMS Herbicides into Meta- and Para-Substituted Series

Weed	*TFMS Series*[b,d]	*0.1% Tween 80 Present*	n[a]	r	r^2	F	SE
Foxtail	3+4+H	No	13	0.820	0.672	6.14	±0.37
	Only 4	No	6	0.978	0.956	14.81	±0.19
	Only 3	No	6	0.992	0.984	92.90	±0.10
Foxtail	3+4+H	Yes	12	0.730	0.533	3.05	±0.45
	Only 4	Yes	6	0.972	0.944	11.17	±0.15
	Only 3	Yes	5	0.999	0.997	127.76	±0.08
Wild Mustard	3+4+H	No	13	0.705	0.498	2.97	±0.41
	Only 4	No	6	0.988	0.976	27.04	±0.16
	Only 3	No	6	0.946	0.896	5.71	±0.20
Wild Mustard	3+4+H	Yes	12	0.776	0.602	4.03	±0.49
	Only 4	Yes	6	0.988	0.976	26.93	±0.19
	Only 3	Yes	5	0.949	0.901	9.04	±0.25
Cheat Grass[c]	3+4+H	No	12	0.602	0.363	1.52	±0.58
	Only 4	No	6	0.990	0.979	31.68	±0.16
	Only 3	No	5	0.997	0.994	176.18	±0.03

[a] 4-SCH_3– and 3-SCH_3–TFMS data omitted from all correlations. *See* following section.

[b] 3-Cl–TFMS data omitted from fits of 3-series in presence of Tween 80 because of experimental difficulties in measuring its partition coefficient (*see* Table VI).

[c] 3-SO_2CH_3–TFMS data omitted from Cheat Grass fits because of difficulty in establishing a precise LD_{90} value for this derivative acting on this weed (*see* Table VII).

[d] 3=meta-substituted TFMS compounds. 4=para-substituted TFMS compounds. H=parent TFMS compound.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Methylthio TFMS Derivatives. Data for the meta- and para-substituted methylthio TFMS derivatives (3-SCH_3, 4-SCH_3) were not included in Hansch structure–activity correlations for the several weed types. This omission was a result of our noting that when herbicidal data for the methylthio derivatives were included in fits, much poorer Hansch correlations were obtained. This was true whether or not the TFMS compounds were separated into meta- and para-derivatives for fitting purposes. It was also true for all weed types examined, in the presence and absence of surfactant. A typical example of the improvement in statistical parameters effected by omitting methylthio points from the data pool is illustrated by the following Hansch correlation of TFMS activity on Foxtail grass in the presence of surfactant.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Foxtail Grass—(0.1% Tween 80 Added)

I. All 3- and 4-Substituted TFMS + Parent

(a) Include 4-SCH_3–, 3-SCH_3–TFMS (n = 14)

$$\log (1/LD_{90}') = \underset{(\pm 1.053)}{-0.571\pi'^2} \underset{(\pm 1.356)}{- 0.976\pi'} \underset{(\pm 0.718)}{+ 1.198\sigma} - 1.572$$

r	r^2	SE	F	
0.575	0.331	±0.53	1.65	(22)

(b) Omit 4-SCH_3–, 3-SCH_3–TFMS (n = 12)

$$\log (1/LD_{90}') = \underset{(\pm 0.924)}{0.310\pi'^2} \underset{(\pm 1.190)}{- 0.720\pi'} \underset{(\pm 0.629)}{+ 1.278\sigma} - 1.690$$

0.730	0.533	±0.45	3.05	(23)

II. Only 4-Substituted TFMS Derivatives

(a) Include 4-SCH_3–TFMS (n = 7)

$$\log (1/LD_{90}') = \underset{(\pm 2.811)}{0.331\pi'^2} \underset{(\pm 3.863)}{+ 0.260\pi'} \underset{(\pm 1.015)}{+ 0.596\sigma} - 1.093$$

r	r^2	SE	F	
0.730	0.533	±0.38	1.14	(24)

(b) Omit 4-SCH_3–TFMS (n = 6)

$$\log (1/LD_{90}') = \underset{(\pm 1.314)}{-2.443\pi'^2} \underset{(\pm 1.830)}{- 3.745\pi'} \underset{(\pm 0.482)}{+ 1.656\sigma} - 2.473$$

0.972	0.944	±0.15	11.17	(25)

III. Only 3-Substituted TFMS Derivatives

(a) Include 3-SCH_3–TFMS ($n = 6$)

$$\log (1/LD_{90}') = \underset{(\pm 2.118)}{-1.544\pi'^2} \underset{(\pm 2.595)}{- 2.111\pi'} \underset{(\pm 2.510)}{+ 3.532\sigma} - 2.503$$

r	r^2	SE	F	
0.744	0.544	±0.85	0.83	(26)

(b) Omit 3-SCH_3–TFMS ($n = 5$)

$$\log (1/LD_{90}') = \underset{(\pm 0.210)}{-0.776\pi'^2} \underset{(\pm 0.257)}{- 1.199\pi'} \underset{(\pm 0.242)}{+ 3.814\sigma} - 2.736$$

0.999	0.997	±0.08	127.76	(27)

In Equations 22–27, note the significant improvements in statistical parameters (r, r^2, F, SE) which arise when 4-SCH_3–TFMS and 3-SCH_3–TFMS data points are removed from the fits. Sequential omission of other single data points (*e.g.*, 3-CF_3–, 3-OH–, 4-CH_3–TFMS, etc.) from the corresponding fit equations produced no such dramatic improvements in statistical parameters, these effects being restricted to the 4-SCH_3– and 3-SCH_3–TFMS derivatives. Further, the example in Equations 24–27 is not atypical. Significant improvements in statistical parameters were also obtained for Foxtail, Wild Mustard, and Cheat Grass in the absence of surfactant when meta- or para-SCH_3 data points were omitted from their respective data pools. In all the above cases, the percent correlation (r^2) was improved from 20–50% upon appropriate deletion of 4-SCH_3– or 3-SCH_3–TFMS data points from the fits. The only two cases where the percent correlation was not very significantly improved by omitting methylthio–TFMS data points were in the fitting of Wild Mustard herbicidal data in the presence of 0.1% Tween 80. In these latter two instances, improvements of only 1–5% in r^2 were obtained by removing 3-SCH_3– and 4-SCH_3–TFMS activity data from their respective data pools. Possible reasons for this relative lack of sensitivity of the final correlation equations for Wild Mustard (in the presence of surfactant) to methylthio–TFMS data inclusion are proposed in the Discussion section. For consistency, however, we have routinely excluded methylthio–TFMS activity data from all the computer determinations of final best fit Hansch relationships discussed in this chapter.

Placement of the Parent Compound. Separation of the TFMS herbicides into two classes consisting of meta- and para-substituted derivatives which were fitted separately resulted in significantly improved correlations *via* the Hansch equation. It was not readily apparent, however, into which of the two herbicidal categories the H-substituted parent com-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

pound should be placed. At first, one might suspect it would fit equally well into either category. The problem of correctly classifying the parent compound was solved using the procedure outlined in Equations 28–31. For the case illustrated below (Foxtail—no Tween 80 present), activity data for the parent compound were first included with the para-substituted series members, and the activity data were refitted to the Hansch equation. The procedure was repeated with the meta-substituted series members. Statistical parameters characterizing the final best-fit correlation equations derived for a data pool containing the TFMS parent compound were then compared with corresponding parameters obtained for correlation equations derived from data not containing the parent compound, and appropriate conclusions were drawn.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Foxtail Grass—(No Tween 80 Present)

4-Substituted TFMS Series Members Alone ($n = 6$)

$$\log (1/LD_{90}) = \underset{(\pm 0.344)}{-1.013\pi^2} + \underset{(\pm 0.179)}{0.332\pi} + \underset{(\pm 0.638)}{3.032\sigma} - 1.079$$

r	r^2	SE	F	
0.978	0.957	±0.19	14.81	(28)

4-Substituted TFMS + Parent Compound ($n = 7$)

$$\log (1/LD_{90}) = \underset{(\pm 0.295)}{-1.108\pi^2} + \underset{(\pm 0.162)}{0.360\pi} + \underset{(\pm 0.566)}{3.164\sigma} - 1.024$$

0.973	0.946	±0.18	17.57	(29)

3-Substituted TFMS Series Members Alone ($n = 6$)

$$\log (1/LD_{90}) = \underset{(\pm 0.068)}{-0.218\pi^2} + \underset{(\pm 0.267)}{3.373\sigma} - 2.060$$

0.992	0.984	±0.10	92.90	(30)

3-Substituted TFMS + Parent Compound ($n = 7$)

$$\log (1/LD_{90}) = \underset{(\pm 0.303)}{-0.229\pi^2} + \underset{(\pm 1.059)}{2.332\sigma} - 1.578$$

0.768	0.589	±0.45	2.86	(31)

Comparison of Equations 28–31 brings to light one important fact. Inclusion of the H-substituted parent compound with the para-substituted series members has little effect on either equational fitting or statistical parameters. In fact, inclusion of the parent with the 4-TFMS compounds

actually results in an improvement in both the statistically important *SE* and *F* parameters. Contrastingly, inclusion of the parent compound with the 3-substituted series members almost destroys the final Hansch structure–activity relationship. For example, equational parameter errors are significantly increased, as is the overall equational standard error *SE*. The *F* value for Equation 31 drops over 30-fold, and the percent correlation (r^2) drops by 40% because of the inclusion of this single parent data point. One must conclude that the parent compound, at least for the illustrated case, does not belong with the meta-TFMS derivatives but fits in satisfactorily with the para-substituted series members.

Comparisons similar to those outlined in Equations 28–31 were likewise undertaken for the remaining cases (Foxtail with Tween 80 present; Wild Mustard with and without Tween 80 present; Cheat Grass without Tween 80 present). The statistical parameters for the final best-fit Hansch correlations for each case are listed in Table IX.

The data in Table IX substantiate the generality of the conclusion reached above—namely, the H-substituted parent TFMS compound behaves herbicidally as though it were a 4- rather than a 3-substituted derivative. For all weed types, in the presence and absence of Tween 80, inclusion of the parent compound with the para-derivatives does not introduce deleterious changes in the statistical parameters. In several of the 4-substituted cases, correlations are improved by its inclusion. In no instance, however, did inclusion of herbicidal activity data for the parent compound with corresponding data for the meta-derivatives result in improved correlations. Except for one case (3-TFMS compounds acting on Wild Mustard in the presence of surfactant), fits were always significantly worsened when the parent compound data were included with those of the 3-derivatives. The lack of a significant parent compound effect for the one exceptional case noted above appears to be related to the fact that for this case LD_{90} exhibits no dependence on the Hammett sigma constant (*see* Table XI and the Discussion for additional details).

On the basis of the statistical data in Table IX, the parent TFMS herbicide is obviously more correctly placed with the 4-substituted series members than with the 3-TFMS compounds. Finally, Table X lists, for comparative purposes, the best-fitting equational and statistical parameters for Hansch correlations of the herbicidal data for 4-substituted series members in which the parent compound has been included. These equations should be compared with those in Table XI.

Optimum Correlation Equations. Table XI gives the final best-fit Hansch relationships which mathematically describe the pre-emergence herbicidal activity of the 3- and 4-substituted TFMS compounds under all conditions examined in this study. The results and conclusions reached

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Table IX. Correlations Including

Weed	*TFMS Series*	*Tween 80 Present*	*Parent TFMS Included*
Foxtail	4-Substituted	No	No
		No	Yes
Foxtail	3-Substituted	No	No
		No	Yes
Foxtail	4-Substituted	Yes	No
		Yes	Yes
Foxtail	3-Substituted	Yes	No
		Yes	Yes
Wild Mustard	4-Substituted	No	No
		No	Yes
Wild Mustard	3-Substituted	No	No
		No	Yes
Wild Mustard	4-Substituted	Yes	No
		Yes	Yes
Wild Mustard	3-Substituted	Yes	No
		Yes	Yes
Cheat Grass	4-Substituted	No	No
		No	Yes
Cheat Grass	3-Substituted	No	No
		No	Yes

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

in the previous sections were incorporated into our fitting procedures—namely:

(a) The TFMS family was divided into two groups consisting of only meta- or only para-substituted derivatives which were fitted separately for each weed type in the presence and absence of Tween 80.

(b) Data for methylthio derivatives (4-SCH_3–TFMS, 3-SCH_3–TFMS) were eliminated from all correlations.

(c) For the equations listed in Table XI, data for the H-substituted parent compound were omitted from all correlations. Comparative equations which include parent-compound data for 4-substituents are given in Table X.

It is evident from the statistical tabulation in Table XI that the TFMS herbicidal data for Foxtail, Cheat Grass, and Wild Mustard are well-correlated by the given Hansch relationships, both in the presence and absence of surfactant. The numbers in parentheses beneath the fit parameter coefficients in Table XI are the standard errors in those coefficients. Perusal of Table XI leads to the following observations:

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Parent TFMS (H-Substituted)

Statistical Parameters for Optimum Hansch Fit

n	r	r^2	SE	F
6	0.978	0.957	±0.19	14.81
7	0.973	0.946	±0.18	17.57
6	0.992	0.984	±0.10	92.90
7	0.768	0.589	±0.45	2.86
6	0.972	0.944	±0.15	11.17
7	0.955	0.913	±0.15	10.46
5	0.999	0.997	±0.08	127.76
6	0.803	0.645	±0.67	1.21
6	0.988	0.976	±0.16	27.04
7	0.988	0.976	±0.13	41.34
6	0.946	0.900	±0.20	5.71
7	0.723	0.522	±0.39	1.09
6	0.988	0.976	±0.19	26.93
7	0.988	0.976	±0.16	40.47
5	0.949	0.900	±0.25	9.04
6	0.909	0.826	±0.28	7.13
6	0.990	0.979	±0.16	31.68
7	0.945	0.894	±0.31	8.41
5	0.997	0.994	±0.03	176.18
6	0.877	0.770	±0.16	5.02

(1) With the exception of the meta-substituted TFMS herbicides acting on Wild Mustard in the presence of Tween 80, the herbicidal activity of all the TFMS compounds exhibits a strong dependence on the Hammett sigma constant for all weed types and surfactant conditions examined. The coefficient of σ is always positive, indicating that (other things being equal) substituents characterized by large, positive σ constants will exhibit enhanced herbicidal activity. Strong electron-withdrawing substituents (*e.g.*, $-SO_2CH_3$, $-NO_2$, etc.) typically have large, positive sigma constants.

(2) Both 3- and 4-substituted TFMS series members exhibit at least some dependence on π, the octanol/water partitioning parameter. As judged by the magnitudes of the π and π^2 parameter coefficients, one must conclude that the herbicidal activity of the 4-substituted sulfonanilides in general depends much more strongly on partitioning events than does the activity of 3-substituted series members both in the presence and absence of surfactant.

(3) In both the Foxtail and Wild Mustard tests, addition of 0.1% Tween 80 to TFMS herbicidal formulations results in an increase in magnitude of the π and π^2 parameter coefficients for both 3- and 4-sub-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

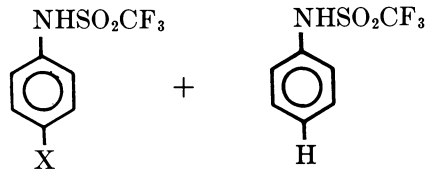
stituted series members. Thus, the addition of small amounts of surfactant to the herbicidal formulations increases the importance of partitioning events in determining overall herbicidal activity.

(4) The π_o values in Table XI are obtained by optimizing the corresponding equations with respect to π, *i.e.*, by taking the derivative of each equation with respect to π (or π'), setting the result equal to zero, and solving for π (or π'). The resulting π_o (or π'_o) value obtained corresponds to either a maximum or minimum on the surface characterizing the accompanying equation. In all cases except one, the π_o value listed corresponds to an optimum value—*i.e.*, the value of π which will give maximum herbicidal activity (at constant σ) to a substituent possessing it. The 3-substituted TFMS compounds acting on Wild Mustard in the presence of Tween 80 show anomalous behavior in that they exhibit a minimum in herbicidal activity over the π range examined.

(5) Referring again to the π_o column in Table XI, note that the addition of surfactant to either 3- or 4-substituted TFMS compounds causes shifts in their herbicidal activity, always toward considerably more negative π values (*see also* Figure 6).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Table X. Final Correlation Equations for 4-Substituted TFMS Herbicides



Fit: $\text{Log}(1/\text{LD}_{90}) = A\pi^2 + B\pi + \rho\sigma + D$ or

Experimental Data						
			Fit Parameters			
Weed Type	*TFMS Series*	*0.1% Tween 80 Present (?)*	A	B	ρ	D
Foxtail	Para	No	−1.108 (±0.295)	0.360 (±0.126)	3.164 (±0.566)	−1.024
	Para	Yes	−2.386 (±1.336)	−3.637 (±1.860)	1.607 (±0.488)	−2.401
Wild Mustard	Para	No	−1.796 (±0.211)	1.275 (±0.116)	3.514 (±0.404)	−0.665
	Para	Yes	−8.400 (±1.372)	−9.851 (±1.910)	3.317 (±0.501)	−3.522
Cheat Grass	Para	No	−1.218 (±0.504)	0.672 (±0.277)	3.686 (±0.966)	−0.998

[a] Data points for 4-SCH_3–TFMS derivatives have been omitted in all correlations (*see* text).

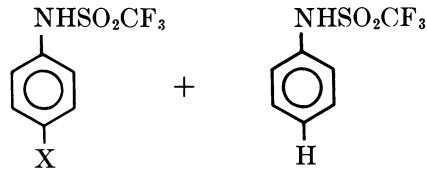
[b] π_o values in the table denote the optimum values of π (or π') producing highest herbicidal activity for a given value of σ.

(6) The π_o values for 3- and 4-substituted series members acting on a given weed type in the absence (or presence) of Tween 80 rarely coincide. In only one instance (Foxtail in the presence of surfactant) did meta- and para-TFMS compounds exhibit the same π-optimum.

A more complete interpretation and comparison of the Hansch relationships in Table XI are given in the Discussion section.

Correlations between Herbicidal Data Obtained in the Presence and Absence of Tween 80. The final correlation equations listed in Table XI for the TFMS herbicides acting on Foxtail, Cheat Grass, and Wild Mustard demonstrate that if the compounds are properly separated into two classes consisting only of meta- or para-substituted series members, the Hansch relationship in its simplest form (Equation 2) can be used to provide good correlations of herbicidal activity with changes in electronic (σ) and partitioning (π) properties of the active molecules. This is true for pre-emergence herbicidal data gathered in the presence and absence of 0.1% Tween 80.

in Which Parent Compound Data are Included (*Cf.* Table XI)[a, b, c]



$$\text{Log}(1/\text{LD}_{90}') = A\pi'^2 + B\pi' + \rho\sigma + D$$

		Statistics		
π_o	n	r	SE	F
+0.16	7	0.973	±0.18	17.57
−0.76	7	0.955	±0.15	10.46
+0.36	7	0.988	±0.13	41.34
−0.59	7	0.988	±0.16	40.47
+0.28	7	0.945	±0.31	8.41

[c] π and LD_{90} values here (*see* Tables VI and VII) are used to fit herbicidal data for tests carried out in the absence of Tween 80. π' and LD_{90}' (*see* Tables VI and VII) values are used to fit data collected in the presence of 0.1% Tween 80. (*See* Table VI and Equation 6 for definition of π and π').

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Table XI. Summary of Correlation

Fit: $\text{Log}(1/\text{LD}_{90}) = A\pi^2 + B\pi + \rho\sigma + D$ or

Experimental Data			*Fit Parameters*			
Weed Type	*TFMS Series*	*0.1% Tween 80*	A	B	ρ	D
Foxtail	Para	No	−1.013 (±0.344)	0.332 (±0.179)	3.032 (±0.638)	−1.079
	Para	Yes	−2.443 (±1.314)	−3.745 (±1.830)	1.656 (±0.482)	−2.473
	Meta	No	−0.218 (±0.068)	0.000 —	3.373 (±0.267)	−2.060
	Meta	Yes	−0.776 (±0.210)	−1.199 (±0.287)	3.814 (±0.242)	−2.736
Wild Mustard	Para	No	−1.794 (±0.278)	1.274 (±0.145)	3.512 (±0.515)	−0.666
	Para	Yes	−8.394 (±1.678)	−9.839 (±2.338)	3.312 (±0.615)	−3.514
	Meta	No	−0.177 (±0.154)	0.371 (±0.104)	1.357 (±0.564)	−1.461
	Meta	Yes	+0.915 (±0.475)	1.930 (±0.615)	0.000 —	−0.516
Cheat Grass	Para	No	−1.532 (±0.280)	0.764 (±0.146)	4.123 (±0.519)	−0.818
	Meta	No	−0.056 (±0.021)	0.000 —	1.854 (±0.106)	−2.011

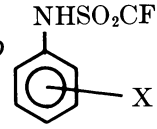
[a] π_o values denote π values producing optimum herbicidal activity at constant σ.
[b] π and LD_{90} values used in correlations of data gathered in the absence of Tween 80. π' and LD_{90}' values similarly used to correlate data gathered in the presence of 0.1% Tween 80 (*see* Tables VI and VII).
[c] π and π' defined in Table VI and Equation *6*. They are both based on $\log P_H = 3.05$ for the unsubstituted parent compound.

In all these previously described correlations, the herbicidal data gathered in the presence and absence of surfactant were always fitted separately. We wished to determine, however, whether activity data gathered in the presence *and* absence of surfactant might be appropriately combined (for a given meta- or para-TFMS classification) to produce a structure–activity relationship of more general applicability. It was of interest, for example, to know whether the LD_{90} values characterizing the action of 4-substituted TFMS compounds on Foxtail (or Wild Mustard) in the absence of Tween 80 could be combined with LD_{90} values characterizing their activity on Foxtail (or Wild Mustard) in the presence of surfactant and the pooled structure–activity data used to derive a Hansch relationship which would describe 4-TFMS activity in either the presence or absence of surfactant. A similar question could be posed for

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Data for TFMS Herbicides[a,b,c]

$$\text{Log}(1/\text{LD}_{90}') = A\pi'^2 + B\pi' + \rho\sigma + D$$



X = meta or para

Statistics

π_o	n	r	SE	F	P[d]
+0.16	6	0.978	±0.19	14.81	93.8
−0.77	6	0.972	±0.15	11.17	91.9
0.00	6	0.992	±0.10	92.90	99.8
−0.77	5	0.999	±0.08	127.76	93.6
+0.36	6	0.988	±0.16	27.04	96.3
−0.59	6	0.988	±0.19	26.93	96.3
+1.05	6	0.946	±0.20	5.71	84.7
−1.05 (minimum)	5	0.949	±0.25	9.04	90.0
+0.25	6	0.990	±0.16	31.68	96.8
0.00	5	0.997	±0.03	176.18	99.4

[d] P is the percentage probability of the F distribution for the corresponding equation (*i.e.*, the probability that a random selection of data points will not be fitted by the listed regression relationship). Values of P corresponding to the F values in the table were obtained by interpolation from a standard F test tabulation (*34*) by plotting percentage probability, P, *vs.* log F (ν_1, ν_2) for the appropriate degrees of freedom ν_1 and ν_2 and then reading the value of P corresponding to our experimental F from the plot.

the 3-substituted series members. Were such a relationship found to be statistically significant, it would imply, for example, that one general Hansch relationship would be sufficient to describe the herbicidal activity of 4-substituted (or 3-substituted) sulfonanilides acting on a given weed type. Addition of surfactant to the system would then not be expected to change the overall shape of the mathematical surface describing biological activity (*i.e.*, the appropriate three-dimensional plot of Equation 2) but rather, would shift the position of the substituted TFMS derivatives on the given surface depending on the effect that surfactant exerted on the partitioning properties (π values) of the compounds in question.

This concept was examined for Foxtail and Wild Mustard by combining herbicidal data gathered in the presence and absence of surfactant for each of the 3- and 4-substituted TFMS series and then carrying out

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Hansch correlations as described above. In performing the correlations, appropriate π and π' values were selected from Table VI; the latter values were used for data gathered in the presence of 0.1% Tween 80. Both π and π' are calculated relative to log $P_H = 3.05$ for the parent compound in the absence of surfactant (*see* Equation 6). Thus, a given TFMS compound in the presence of Tween 80 might be thought of as a new member of the similarly substituted TFMS series characterized by somewhat different partitioning properties than it originally exhibited in the

Table XII. Surfactant and Non-Surfactant

$$\text{Log}(1/\text{LD}_{90}) = A\pi^2 +$$

Experimental Data						
			Fit Parameters			
Weed Type	*TFMS Series*	*0.1% Tween 80 Present (?)*	A	B	ρ	D
Foxtail	Para	No	−1.013 (±0.344)	0.332 (±0.179)	3.032 (±0.638)	−1.079
		Yes	−2.443 (±1.314)	−3.745 (±1.830)	1.656 (±0.482)	−2.473
		Both[a]	−0.022 (±0.194)	−0.043 (±0.125)	0.886 (±0.276)	−1.298
Foxtail	Meta	No	−0.218 (±0.068)	0.00 —	3.373 (±0.267)	−2.060
		Yes	−0.776 (±0.210)	−1.199 (±0.257)	3.814 (±0.242)	−2.736
		Both[a]	−0.024 (±0.109)	−0.058 (±0.084)	3.109 (±0.396)	−2.198
Wild Mustard	Para	No	−1.794 (±0.278)	1.274 (±0.145)	3.512 (±0.515)	−0.666
		Yes	−8.394 (±1.678)	−9.839 (±2.338)	3.312 (±0.615)	−3.514
		Both[a]	−0.752 (±0.408)	0.637 (±0.262)	0.915 (±0.579)	−0.618
Wild Mustard	Meta	No	−0.177 (±0.154)	+0.371 (±0.104)	+1.357 (±0.564)	−1.461
		Yes	0.915 (±0.475)	1.930 (±0.615)	0.000 —	−0.516
		Both[a]	−0.252 (±0.144)	0.361 (±0.110)	1.207 (±0.521)	−1.246

[a] For these correlations, TFMS herbicidal data obtained in the presence and absence of Tween 80 for a given species were combined and fitted simultaneously. π and π' values in Table VI were used. Parent compound data were omitted. *See* text for details.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

absence of surfactant. Table XII presents the equation parameters and statistical results obtained using the above-described fitting procedure.

The results in Table XII show that Hansch correlations carried out on pooled herbicidal data gathered in the presence and absence of Tween 80 for a given classification of TFMS compounds acting on a particular weed type produce significantly poorer ultimate structure–activity relationships than those which are obtained if the activity data gathered in the presence and absence of surfactant are fitted separately. For both

TFMS Data Combined

$B\pi + \rho\sigma + D$

		Statistics		
π_o	n	r	SE	F
+0.16	6	0.978	±0.19	14.81
−0.77	6	0.972	±0.15	11.17
−0.99	12	0.859	±0.29	7.50
0.00	6	0.992	±0.10	92.96
−0.77	5	0.999	±0.08	127.76
−1.20	11	0.958	±0.23	26.12
+0.36	6	0.988	±0.16	27.04
−0.59	6	0.988	±0.19	26.93
+0.42	12	0.677	±0.61	2.26
+1.05	6	0.946	±0.20	5.71
−1.05 (minimum)	5	0.949	±0.25	9.04
+0.72	11	0.831	±0.30	5.20

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

meta- and para-substituted TFMS series, correlations based on the pooled data always exhibited higher parameter standard errors, significantly lower multiple correlation coefficients (r), lower percent correlations (r^2), higher equational standard errors (SE), and lower F ratios than the original correlation equations based on separated data measured in the presence and absence of surfactant. For the most meaningful results, therefore, for both the Foxtail and Wild Mustard cases, data gathered in the presence and absence of surfactant (for both meta- and para-substituted TFMS series) should be fitted separately and not combined. [In a manner similar to that outlined in Table VIII and its accompanying text, Hansch stepwise regression analyses were also carried out separately for Foxtail and Wild Mustard in which herbicidal data points (LD_{90} values) for all TFMS series members (3- + 4- + H-substituted) gathered in the presence and absence of Tween 80 were combined into one data pool and fitted simultaneously. As in Table VIII, very poor structure–activity correlations of the combined data were obtained. The results again demonstrated that the TFMS herbicides must be separated into two families of only meta- or para-substituted derivatives if meaningful structure–activity relationships are to be achieved.] Furthermore, with reference to the question posed earlier in this section, the data in Table XII strongly suggest that the addition of Tween 80 to either the meta- or para-substituted TFMS series results more in a change in shape of the mathematical surface characterizing the Hansch equation appropriate to that series than in large shifts of individual TFMS data points on a common mathematical activity surface (*see also* Discussion). Were this not the case, poorer statistical correlations would *not* have been obtained when the activity data collected in the presence and absence of Tween 80 were pooled (*see* Table XII).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Discussion

Two general questions relating to the pre-emergence activity of the TFMS herbicides have been raised—namely:

(a) Can the Hansch correlation method be used quantitatively to assess the effects of relatively low levels of a formulation surfactant, Tween 80, on TFMS herbicidal activity?

(b) Can the Hansch correlation method be used quantitatively to account for differences in grass and broadleaf activity among TFMS series members?

In view of the data presented, both questions can now be answered in the affirmative. The best-fit Hansch relationships in Table XI can be used to predict and to explain variations in TFMS grass and broadleaf activity. Furthermore, the appropriate equations in Table XI can be used

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

to predict the effects that 0.1% Tween 80 will have on the Wild Mustard and Foxtail activity of TFMS series members.

Besides providing mathematical relationships correlating structural changes to herbicidal activity variations among TFMS series members, Hansch analyses brought to light other facts which undoubtedly would have been overlooked had the TFMS herbicides not been examined by the Hansch method. In particular, the stepwise regression procedure demonstrated that:

(a) The 3- and 4-substituted trifluoromethanesulfonanilides function as two separate families of herbicides, each characterized by its own specific correlation equations. The π and σ dependencies exhibited by the meta- and para-substituted categories are quite different (*see* Table XI), and these differences are maintained in both the presence and absence of surfactant.

(b) Herbicidal activity data for the TFMS methylthio derivatives (4-SCH_3–, 3-SCH_3–) cannot be well correlated using the Hansch relationships (Table XI) which satisfactorily explain the activity of other series members. This suggests either a different mode of action for the methylthio-substituted TFMS series members or the possible conversion of the methylthio side chain to another species with differing reactivity during the course of the herbicidal tests.

(c) Although the H-substituted parent compound is symmetrically substituted about the meta- and para-ring positions, it functions herbicidally as though it were a 4-substituted series member.

(d) Separate Hansch relationships are required to characterize the herbicidal activity of the 4-substituted TFMS compounds in the presence and absence of Tween 80. A similar statement applies to the 3-substituted series members.

Perspective Plots of TFMS Herbicidal Activity Surfaces. The final correlation equations in Table XI quantitatively account for the observed herbicidal activity differences between 3- and 4-substituted TFMS derivatives in the presence and absence of surfactant. These differences manifest themselves as variations in the signs and magnitudes of the π^2, π, and σ parameters in the Hansch relationships. Although much information is contained in these parameter variations, the overall complexity of the mathematical expressions makes it difficult to appreciate fully their implications. To help us visualize the significant activity differences of the 3- and 4-substituted TFMS herbicides, the computer has been used to construct three-dimensional perspective plots of the surfaces represented by the best-fit Hansch relationships in Table XI. These surfaces are illustrated in Figures 2a–2d for Foxtail, Figures 3a–3d for Wild Mustard, and Figures 4a–4b for Cheat Grass. In order that differences between plots of the several biological activity surfaces be more easily visualized, the usual Cartesian coordinate system has been rotated from its normal standard position as shown in Figure 5. The final positions

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

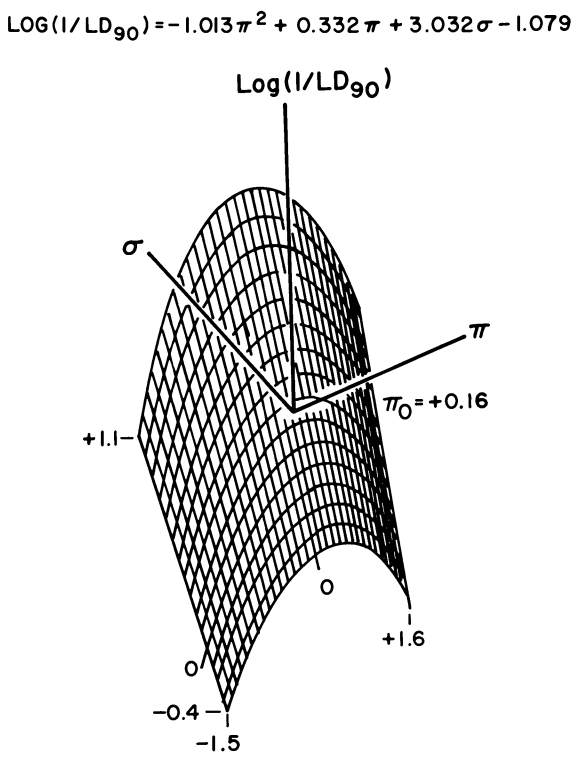


Figure 2a. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 4-substituted TFMS herbicides on Foxtail grass (no Tween 80 present)

of the respective axes should be firmly set in mind at this point so that there will be no subsequent confusion in interpreting the plots.

In the final position, the positive end of the $\log(1/LD_{90})$ axis is tipped 45° out of the plane of the page toward the viewer and the negative end 45° into the plane of the page. All of the plots of Figures 2, 3, and 4 are identically positioned in the final orientation shown in Figure 5. The respective ranges of π and σ on all plots are likewise identical. The range of π in all plots is -1.5 to $+1.6$ while σ always ranges from -0.4 to $+1.1$. The increments on all plots are in units of 0.1 in both the π and σ directions. Thus, even though the axes themselves are not divided into increments, the position of any given (π, σ) point can be readily located on any particular biological activity surface. The value of $\log(1/LD_{90})$ corresponding to any (π, σ) point (*i.e.*, the vertical height of the activity surface) can be calculated quantitatively for any figure through use of

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

the appropriate Hansch relationships in Table XI. The ranges of π and σ in the plots were chosen to overspan slightly the ranges of our experimental π's and σ's in Tables IV and V. Thus, the plots are not grossly extrapolated outside the range of our experimental data points and should be useful for predictive purposes.

Since perspective plots like those in Figures 2–4 are subject to optical illusions, instruction as to their proper viewing is necessary. The parabolic surfaces of Figures 2a, 2c, 2d, 3a, 3c, and 4a all open downward. The viewer looks at the top surface of the plot rather than at the underside. Similarly, in the more planar plots of Figures 2b, 3b, and 4b the view is of the top surface of the plots. Figure 3d, which exhibits a minimum in the activity surface (*see* Table XI), rises to the right. The view is again of the top surface.

Except for the Hansch relationship of Figure 3d which exhibits no dependence on σ, all other Hansch relationships in Table XI have a fairly large, positive σ coefficient. Hence, the value of the $\rho\sigma$ term will increase in a positive fashion with increasing σ in all of these latter equations. With respect to the corresponding plots of Figures 2–4, this means that

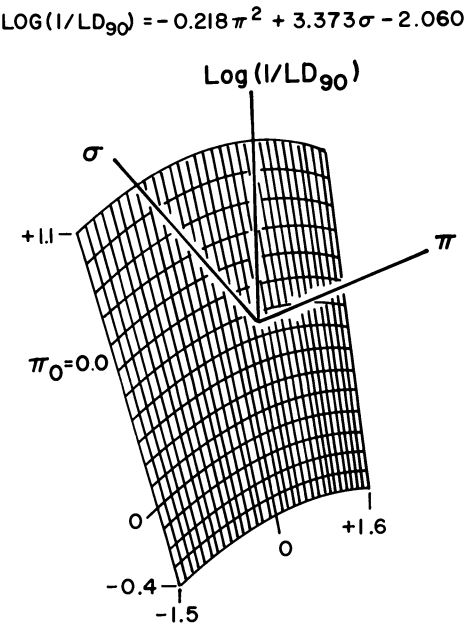


Figure 2b. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 3-substituted TFMS herbicides on Foxtail grass (no Tween 80 present)

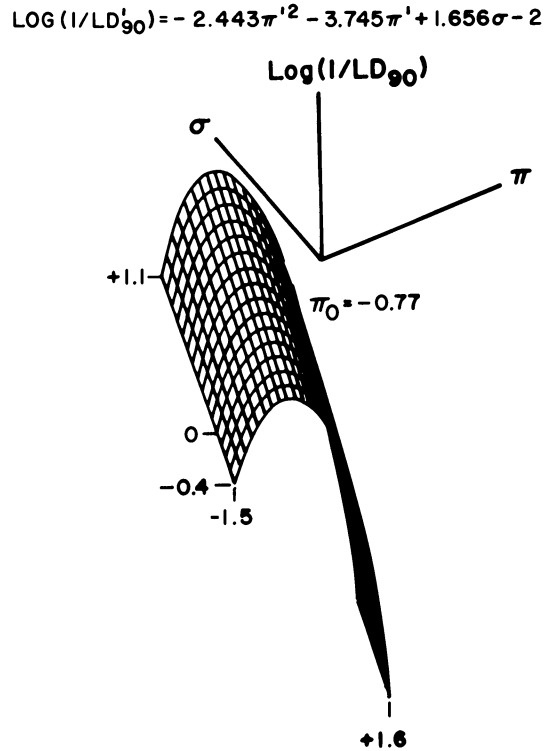


Figure 2c. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 4-substituted TFMS herbicides on Foxtail grass (0.1% Tween 80 present)

in most cases (depending on the magnitude of the σ coefficient, ρ), the surface rises rather sharply away from the viewer in the positive σ direction. Keeping in mind that the negative $\log(1/LD_{90})$ axis is tilted 45° into the plane of the page, the actual rise of the herbicidal activity surfaces in all cases is somewhat less steep than it appears in the plots.

The origin is indicated on all plots of Figures 2–4 as the intersection point of the π, σ, and $\log(1/LD_{90})$ axes. All surfaces are plotted relative to this origin. Since scales in the π, σ, and $\log(1/LD_{90})$ directions are identical in all figures, the respective plotted surfaces are all directly comparable in terms of their displacement from this origin. If it is feasible to make copies or transparencies of Figures 2 to 4, the various plots can be visually compared easily by overlapping two or more of them, aligning the corresponding axes, and viewing the stack either by placing them on a light box or by holding them up toward a light source.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

TFMS Herbicides in the Absence of Tween 80. Figures 2a, 3a, and 4a depict the herbicidal action of 4-substituted TFMS series members on Foxtail, Wild Mustard, and Cheat Grass, respectively, in the absence of 0.1% Tween 80. Although all the surfaces demonstrate a generally similar parabolic dependence on π, they exhibit differing π optima (π_o values, *see* Table XI). The positions of the π optima on the above-cited plots can be accurately located by counting an appropriate number of units from the origin along the positive π axis in each figure (recall that each division = 0.1 π unit). The main difference in the plots lies in the steepness of their rise in the positive σ direction. The Cheat Grass plot in Figure 4a exhibits the steepest rise because of the large magnitude of its σ coefficient (ρ = 4.12, *see* appropriate Hansch relationship in Table XI). The most active 4-substituted TFMS derivatives for all grass types will be those characterized by π values lying on the π optimum and σ values as large and positive as possible. Derivatives whose final position on the activity surface lies in the π–σ plane (which includes the origin) will be very active since in this plane, $\log(1/LD_{90})$ = 0.0 and LD_{90} = 1.0 mole/acre.

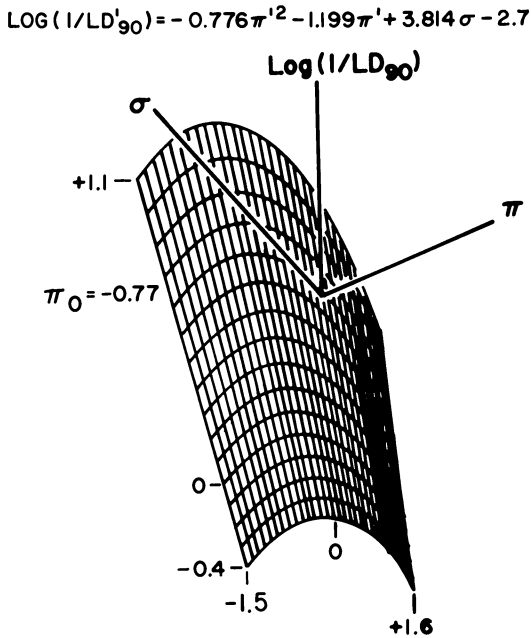


Figure 2d. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 3-substituted TFMS herbicides on Foxtail grass (0.1% Tween 80 present)

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Comparing now the corresponding plots for 3-substituted TFMS compounds on these same weeds in the absence of surfactant (Figures 2b, 3b, and 4b), we note a surprising difference in activity surface shapes relative to the 4-substituted series members. Although the surfaces of the above figures all have a slight parabolic π dependence (*cf.* A and B parameter coefficients in Table XI), the plots on the whole are almost planar relative to those of the corresponding 4-substituted TFMS derivatives. The Foxtail plot of Figure 2b exhibits the steepest rise in the positive σ direction, again because of the comparatively large value of $\rho = 3.37$ for this particular surface (*cf.* Table XI).

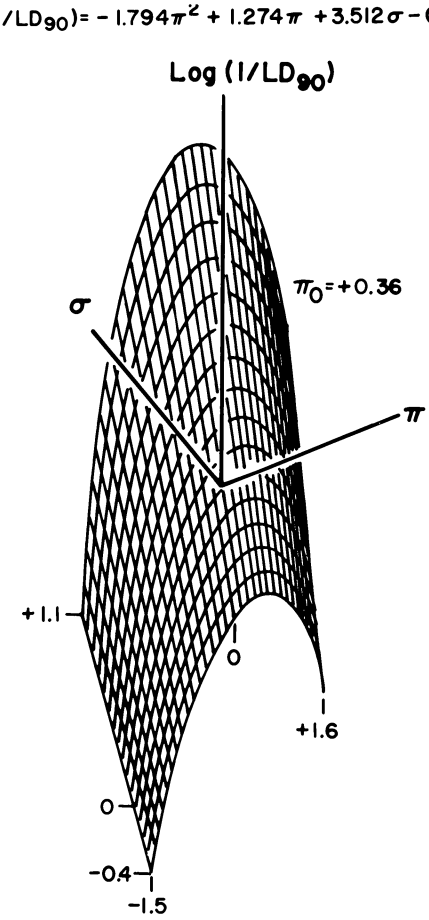


Figure 3a. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 4-substituted TFMS herbicides on Wild Mustard (no Tween 80 present)

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

$$LOG(1/LD_{90}) = -0.177\pi^2 + 0.371\pi + 1.357\sigma - 1.461$$

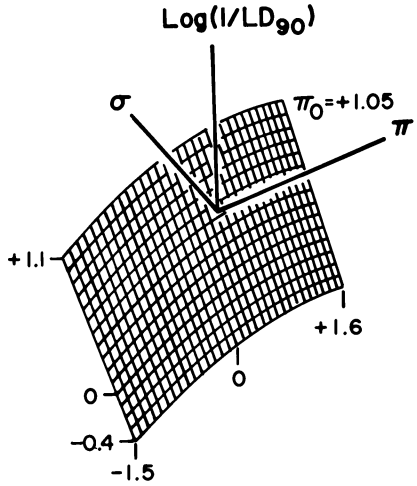


Figure 3b. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 3-substituted TFMS herbicides on Wild Mustard (no Tween 80 present)

The plots of Figures 2a–b, 3a–b, and 4a–b show why poor Hansch correlations were obtained for herbicidal data unless the TFMS compounds were first separated into two groups consisting only of 3- or 4-substituted derivatives. Since the surfaces characterizing the activity of the 4-substituted derivatives exhibit such a strong parabolic dependence on π relative to the 3-TFMS compounds, it is evident that for a given weed type, only at the points of intersection of the meta- and para-activity surfaces will a single Hansch relationship accurately fit the data. Data points removed from this locus of intersection will be poorly fitted by a common regression relationship. Alternatively, data points lying on the locus of intersection will be equally well fitted by either the appropriate meta- or para-Hansch relationships in Table XI. This fact brings up an important general point. If the TFMS derivatives had been arbitrarily chosen for evaluation such that their π and σ coordinates were on or near the locus of intersection of the meta and para surfaces for a given weed type, Hansch analyses would have predicted a single final regression equation which would have fitted both 3- and 4-substituted TFMS derivatives equally well. One could thus have concluded wrongly that meta- and para-substituted compounds behave herbicidally in a similar fashion. π and σ values characterizing the substituents of a family of biologically active materials should thus always be chosen to span as broad

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

a range of these parameters as is practical to avoid potential difficulties of the above type.

TFMS Herbicides in the Presence of 0.1% Tween 80. Let us now turn our attention to the effect that the addition of a small amount of Tween 80 to the TFMS herbicidal formulations has on the shape and orientation of the activity plots for Foxtail and Wild Mustard. Considering first the 4-substituted derivatives and comparing Figure 2a with 2c for Foxtail and Figure 3a with 3c for Wild Mustard, we note that the addition of surfactant in both cases has two obvious consequences:

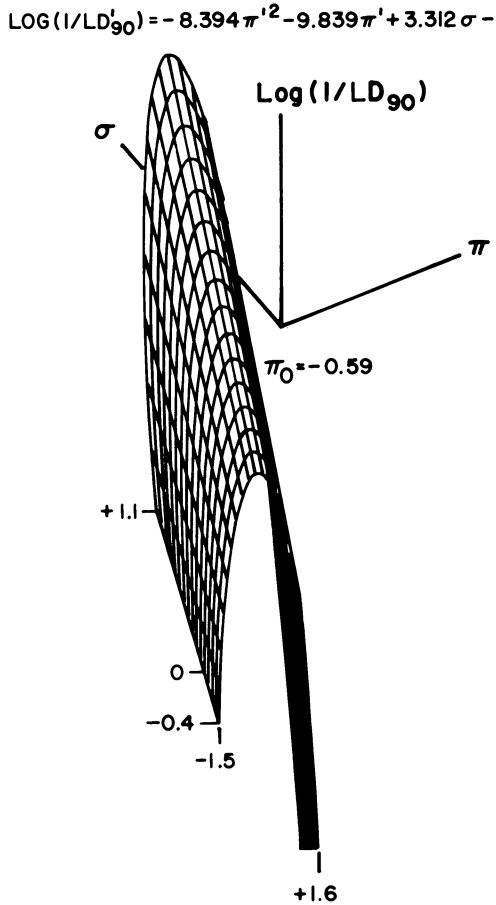


Figure 3c. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 4-substituted TFMS herbicides on Wild Mustard (0.1% Tween 80 present)

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

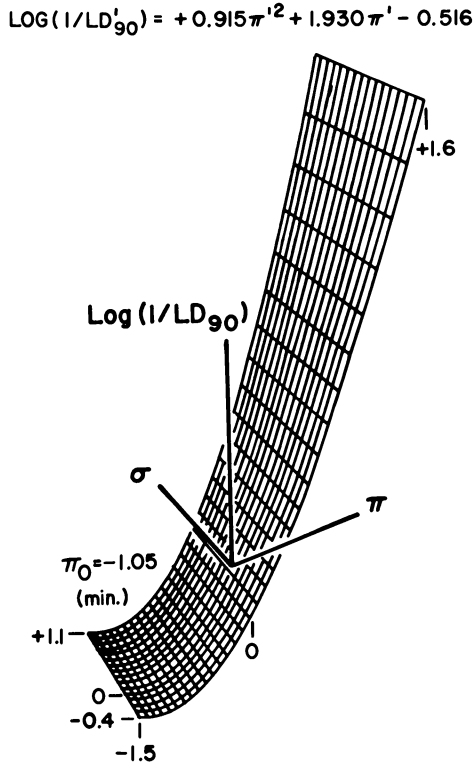


Figure 3d. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 3-substituted TFMS herbicides on Wild Mustard (0.1% Tween 80 present)

(a) It shifts the π-optimum value to significantly more negative π values.

(b) It sharply decreases the broadness of the parabolic surface characterizing 4-TFMS activity on Foxtail and Wild Mustard.

In practical terms, this means that in the presence of surfactant, one has far less latitude in choosing a substituent with a π value near the π optimum than one does in the absence of surfactant. For example, in the presence of Tween 80, small deviations in π from π_o (relative to those permissible in the absence of surfactant) will lead to a rapid decrease in herbicidal activity. The loss in 4-TFMS activity produced by such deviations should be particularly severe for Wild Mustard, which has such a steep parabolic dependence on π in the presence of Tween 80 (*see* Figure 3c).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In addition to the pronounced effects of surfactant on π discussed above, addition of Tween 80 to 4-TFMS herbicidal formulations produces smaller changes in the σ dependence (*cf.* Figures 2a and 2c for Foxtail and 3a and 3c for Wild Mustard). In Table XI, this σ dependence appears as a slight decrease in the parameter coefficient ρ for 4-TFMS activity on Wild Mustard and as a large decrease in ρ for 4-TFMS activity on Foxtail when surfactant is added. Comparing Figures 2a and 2c for Foxtail, this decrease in ρ manifests itself as a decrease in slope of the activity surface in the positive σ direction. For substituents with large positive σ values,

$$\text{LOG}(1/LD_{90}) = -1.532\pi^2 + 0.764\pi + 4.123\sigma - 0.818$$

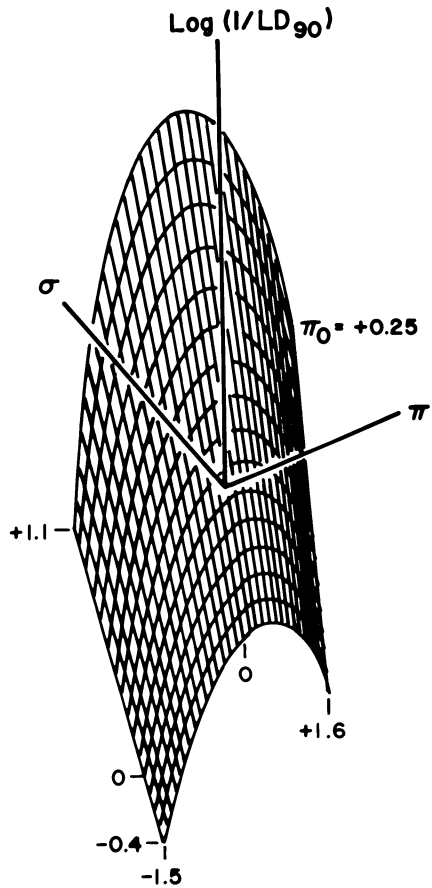


Figure 4a. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 4-substituted TFMS herbicides on Cheat Grass (no Tween 80 present)

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

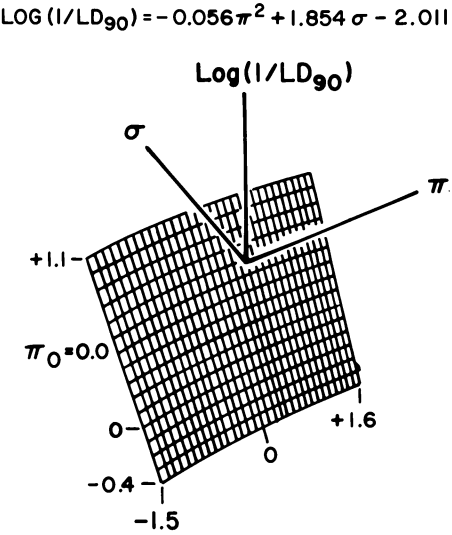


Figure 4b. Three-dimensional perspective plot of the Hansch equation describing pre-emergence activity of 3-substituted TFMS herbicides on Cheat Grass (no Tween 80 present)

the plots indicate that herbicidal activity will be depressed when surfactant is added, particularly if π for the substituent lies relatively near the optimum π_o value in the presence and absence of surfactant. The effect of surfactant on σ is much less pronounced in the Wild Mustard plot of Figure 3c because of the considerably smaller decrease in ρ when surfactant is added (*see* Table XI).

With reference to Figures 2a for Foxtail and 3a for Wild Mustard, note that the left sides of both parabolic activity surfaces (corresponding to hydrophilically substituted TFMS derivatives) are shifted only slightly towards more negative π values when Tween 80 is added (*see* Figures 2c and 3c). In contrast, their right parabolic surfaces (corresponding to lipophilically substituted series members) are shifted drastically towards more negative π values. As pointed out in the Results section, the shift in optimum π_o (coupled with a sharpening of the respective activity surfaces) comes about largely because of the significant water-solubilizing effect of Tween 80 on hydrophobic TFMS substituents and its small effect on hydrophilic substituents.

The effect of adding Tween 80 to herbicidal formulations of 3-substituted TFMS compounds is likewise dramatic, as evidenced by Figures 2b and 2d for Foxtail and 3b and 3d for Wild Mustard. For the Foxtail case, addition of surfactant shifts the π optimum for 3-substituents from

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

$\pi_o = 0.0$ to $\pi_o = -0.77$ and introduces more partitioning dependence into the activity surface (note also the increase in A and B coefficients for the appropriate equations in Table XI when Tween 80 is added). Figure 2d shows that the herbicidal activity of the more hydrophobically substituted 3-TFMS series members (larger positive π values) will tend to be inhibited by surfactant addition. This inhibitory effect manifests itself as a downward bending of the activity surface on the hydrophobic side relative to its position in the absence of surfactant (*cf.* Figures 2b and 2d).

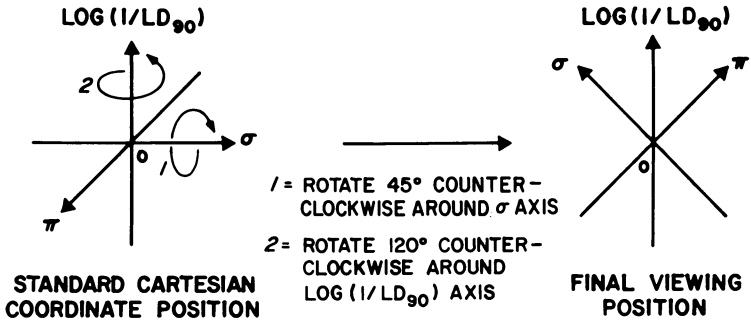


Figure 5. Rotation of coordinate axes

3-TFMS herbicidal activity on Wild Mustard is also affected in an interesting way by the addition of Tween 80. The fairly flat, planar, almost π-independent surface of Figure 3b is converted in Figure 3d to a highly π-dependent curve which exhibits rapidly increasing herbicidal activity with increasing π. This behavior is in complete contrast to that observed in any other case. Also, addition of surfactant completely eliminates the σ dependence of 3-TFMS activity on Wild Mustard ($\rho = 0.0$, *see* Table XI). In contrast, the σ dependence of 3-substituted sulfonanilides on Foxtail is slightly enhanced by the addition of surfactant [$\rho =$ 3.37 (no Tween 80); $\rho = 3.81$ (with Tween 80)].

From Table XI, note that the π optimum for both 3- and 4-substituted TFMS series members acting on Foxtail grass in the presence of surfactant is $\pi_o = -0.77$. Examination of Figures 2c and 2d indicates, however, that gross differences in the herbicidal activity surfaces for the meta- and para-TFMS series exist. On this basis, it would appear that coincidence of π_o values for 3- and 4-substituted series is not a sufficient condition for assuming that modes of herbicidal action for the two series are also identical.

One point should be clarified with reference to the plots of Figures 2, 3, and 4 and the Hansch relationships in Table XI. The appropriate equations and plots describing TFMS herbicidal activity in the absence

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

of surfactant are all based on the π values of Table VI. Equations and plots describing activity in the presence of Tween 80 are correspondingly based on the π' values in the same table. As previously explained, π and π' are calculated relative to log $P_H = 3.05$ for the parent compound in the absence of surfactant. If one instead wishes to choose as a reference point the parent compound in the presence of 0.1% Tween 80, the π'' values in Table VI should be used in the appropriate regression analyses. With respect to the equations of Table XI and the plots of Figures 2–4, the use of π'' instead of π' values in fitting the herbicidal data obtained in the presence of surfactant has no effect other than to shift the origin of the coordinate axes -0.45 π units in the negative π direction (*i.e.*, to the position of the parent compound in the presence of 0.1% Tween 80). Neither the shape nor the vertical position of the surfaces of Figures 2c–d or 3c–d would be affected in any way by such a switch to π'' values from π' values. Hence, all conclusions previously reached by using π' values apply equally well to π'' values.

Finally, Table XII showed that for both 3- and 4-substituted TFMS series, poorer correlations were obtained when herbicidal data obtained in the presence and absence of Tween 80 for a given series acting on a particular weed type were pooled and fitted simultaneously. Comparing the appropriate surface plots in Figures 2–3 for Foxtail and Wild Mustard, it is evident that poor correlations of the pooled data would have been obtained. The very different graphical curve shapes obtained for each TFMS series in the presence and absence of Tween 80 illustrate why correlations of the pooled data produced poor structure–activity correlations.

Predictive Capabilities of Hansch Correlation Relationships. The true test of any mathematical structure–activity relationship is whether it can be used: (a) to account for observed differences in biological activity of series members used in deriving the correlation equations, and (b) to predict the activity of other series members not included in the data fitting pool, or alternately, predict the activity of new optimum activity series members not yet investigated.

As an illustration of the use of the final Hansch relationships in Table XI in explaining biological activity differences between individual TFMS series members, we will examine the activity of 4-SO_2CH_3–TFMS and 4-Cl–TFMS against Foxtail, Wild Mustard, and Cheat Grass. Table XIII gives values of the individual terms in the Hansch equations of Table XI which are appropriate to the action of each herbicide on the three weed types. At a later point, we also examine the activity of 2,4-di-Cl– and 2,4-di-F–TFMS derivatives as predicted by the Hansch relationships in Table XI.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Table XIII. Herbicidal Activity of 4-SO_2CH_3–TFMS

Weed Type	Tween 80 Present (?)	Value of Hansch Equation Terms: $A\pi^2$	$B\pi$	$\rho\sigma$	D	log $(1/LD_{90})$
4-SO_2CH_3–TFMS ($\pi = -1.06$, $\pi' = -1.25$, $\sigma = +0.728$)						
Foxtail	No	−1.138	−0.352	+2.207	−1.079	−0.362
	Yes	−3.817	+4.681	+1.206	−2.473	−0.403
Wild Mustard	No	−2.016	−1.350	+2.557	−0.666	−1.475
	Yes	−13.116	+12.299	+2.411	−3.514	−1.920
Cheat Grass	No	−1.721	−0.810	+3.002	−0.818	−0.348
4-Cl–TFMS ($\pi = +0.91$, $\pi' = -0.45$, $\sigma = +0.227$)						
Foxtail	No	−0.839	+0.302	+0.688	−1.079	−0.927
	Yes	−0.495	+1.685	+0.376	−2.473	−0.907
Wild Mustard	No	−1.486	+1.159	+0.797	−0.666	−0.195
	Yes	−1.700	+4.428	+0.752	−3.514	−0.034
Cheat Grass	No	−1.269	+0.695	+0.936	−0.818	−0.455

4-SO_2CH_3–TFMS. Considering first the 4-SO_2CH_3–TFMS derivative, we note from the LD_{90} tabulation (Table VII) that it exhibits poor broadleaf (Wild Mustard) herbicidal activity but good grass (Foxtail, Cheat Grass) activity. Surfactant has little effect on 4-SO_2CH_3–TFMS Foxtail activity but significantly inhibits its Wild Mustard activity. The LD_{90} values tabulated in Table XIII and calculated with the appropriate Hansch relationships in Table XI are in excellent agreement with the experimental LD_{90} values listed in Table VII. As shown in Table XIII, our derived structure–activity relationships correctly predict that the 4-SO_2CH_3–TFMS derivative will have good grass and poor broadleaf herbicidal activity, as is observed experimentally. They also predict that Tween 80 will have little effect on the Foxtail activity of 4-SO_2CH_3–TFMS but will greatly inhibit its Wild Mustard activity.

Much explanatory information is also contained in the individual terms of the several Hansch relationships in Table XIII. For example, even though the Foxtail activity of 4-SO_2CH_3–TFMS appears to be unaffected by Tween 80 (as judged by the small effect of surfactant on LD_{90}), examination of each term in the Hansch equations for Foxtail in Table XIII reveals that Tween 80 does indeed affect activity, but in a subtle way. Comparing the $\rho\sigma$ terms for Foxtail in the presence and absence of surfactant, Tween 80 greatly reduces the magnitude of this term. Taken by itself, this surfactant effect on $\rho\sigma$ would have considerably reduced 4-SO_2CH_3–TFMS Foxtail activity. Tween 80, however, also shifts the π_o optimum for Foxtail activity to a more negative value, which is much closer to the π' value of −1.25 characterizing the 4-SO_2CH_3–

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

and 4-Cl–TFMS on Various Weed Types

Calculated LD_{90}	*Experimental* LD_{90} *from Table VII*
2.30	2.42
2.53	2.64
29.88	31.6
83.13	79.0
2.23	2.21
8.46	6.40
8.06	8.15
1.57	1.36
1.08	1.35
2.85	1.95

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

TFMS derivative in the presence of surfactant (*cf.* also Figures 2a and 2c). Thus, the loss in Foxtail activity which results from the deleterious effect of surfactant on the $\rho\sigma$ term in the Hansch equation is compensated by the improved partitioning characteristics (large $B\pi$ term) of the 4-methylsulfonyl derivative in the presence of surfactant. The net result is that the overall pre-emergence Foxtail activity is not greatly affected by Tween 80. As illustrated in Table XIII, however, this effect is only apparent since significant changes in both π and σ dependence have taken place.

The effect of surfactant on 4-SO_2CH_3–TFMS Wild Mustard activity can be treated in a similar manner. From Table XIII note that the $\rho\sigma$ term is relatively unchanged by the addition of surfactant. The observed significant drop in activity results from a combination of poor partitioning properties ($A\pi^2 + B\pi$ negative) and a general translation of the herbicidal activity surface in the negative $\log(1/LD_{90})$ direction (note the large negative intercept value D). These effects can be observed by comparing Figures 3a and 3c.

4-SO_2CH_3–TFMS is even more active on Cheat Grass than it is on Foxtail despite the fact that π (4-SO_2CH_3–) $= -1.06$ is even further removed from the π_o optimum for Cheat Grass ($\pi_o = +0.25$, *see* Table XI) than it is for Foxtail ($\pi_o = +0.16$). From Table XIII (and Figure 4a), we see that the high Cheat Grass activity is entirely caused by the favorable value of the $\rho\sigma$ term in the appropriate Hansch relationship. In this case, both $\rho = 4.123$ and $\sigma = +0.728$ are favorably large and positive.

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

4-Cl–TFMS. From Table VII note that 4-Cl–TFMS is considerably more active on the broadleaf Wild Mustard than it is on Foxtail grass. As shown in Table XIII, the appropriate Hansch relationships in Table XI also predict this difference in grass and broadleaf activity. Term-by-term analysis of the several appropriate correlation equations demonstrates that for Foxtail in the absence of surfactant, non-optimum partitioning characteristics (π (4-Cl) = +0.91; π_o = +0.16) and a relatively low $\rho\sigma$ value both contribute to the diminished grass herbicidal activity of 4-Cl–TFMS. Although partitioning characteristics of this derivative are improved by Tween 80 addition (more positive $A\pi^2 + B\pi$ contribution; π'(4-Cl) = −0.45 close to π_o' = −0.77), the $\rho\sigma$ contribution is diminished because of a significant decrease in ρ caused by surfactant (*see* Table XI). In addition, Foxtail activity is depressed to some extent in the presence of surfactant because of a translational shift of the activity surface in the negative log(1/LD_{90}) direction (*cf.* Figures 2a and 2c). Hence, overall Foxtail activity is reduced in the presence of surfactant.

Wild Mustard activity of the 4-Cl–TFMS derivative in both the presence and absence of Tween 80 is higher than that observed for Foxtail because both π = +0.91 and π' = −0.45 for this compound lie closer to their appropriate π optima (π_o = +0.36 (no surfactant); π_o' = −0.59 (with surfactant)) than was the case for Foxtail. Hence, the $A\pi^2 + B\pi$ contribution for the broadleaf in both the presence and absence of surfactant is more favorable than it is for the grass. The $\rho\sigma$ contribution is also more favorable for Wild Mustard because of the high value of ρ for 4-substituted TFMS compounds in the presence and absence of Tween 80 (*see* Table XI). Thus, for the 4-Cl–TFMS derivative, both enhanced partitioning characteristics and a more favorable electronic situation (more positive $\rho\sigma$) combine to give higher Wild Mustard than Foxtail activity.

Cheat Grass activity in the absence of surfactant for the 4-Cl–TFMS derivative is again higher than corresponding Foxtail activity primarily because of the larger positive value of the $\rho\sigma$ term. π(4-Cl) = +0.91 also lies slightly closer to the π optimum for Cheat Grass activity (π_o (Cheat Grass) = +0.25) than it does to the π optimum for Foxtail activity (π_o (Foxtail) = +0.16).

Thus, term-by-term analysis of the Hansch relationships derived for each TFMS series in the presence and absence of Tween 80 (coupled with perspective plots of the corresponding herbicidal activity surfaces) permits one to account for activity differences between species in a fairly detailed though empirical manner. Indeed, whereas Trepka and coworkers (5) found it difficult to account for grass and broadleaf selectivity differences of the 4-Cl– and 4-SO_2CH_3–TFMS derivatives in a straight-

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

forward manner (*see* Table I), the preceding treatment goes a long way toward both explaining and predicting these differences.

2,4-Di-Cl–TFMS AND 2,4-Di-F–TFMS. Trepka *et al.* (*5*) found the 2,4-di-Cl– and 2,4-di-F–TFMS derivatives to be highly active against both grasses and broadleaves (*see* Table I). We may thus ask whether the appropriate Hansch relationships listed in Table XI for the mono-substituted TFMS derivatives can be used to predict the high activity of these disubstituted compounds.

In their studies of multisubstituted aromatic compounds, Hansch and co-workers have generally summed π and σ values for individual substituents to obtain estimates of π and σ characterizing the multisubstituted derivative (*10*). Although π and σ constants for ortho substituents may not be well represented by the corresponding values for para substituents, let us assume that π(2-Cl) = π(4-Cl) and σ(2-Cl) = σ(4-Cl) for our predictive calculation. For the disubstituted TFMS derivatives, we obtain from Table IV (in the absence of surfactant):

2,4-di-Cl-TFMS	*2,4-di-F-TFMS*
$\Sigma\pi = 1.82$	$\Sigma\pi = +0.40$
$\Sigma\sigma = +0.454$	$\Sigma\sigma = +0.124$

Using these $\Sigma\pi$ and $\Sigma\sigma$ values in conjunction with the first, fifth, and ninth para Hansch relationships in Table XI, the grass and broadleaf activity in the absence of Tween 80 for the two disubstituted derivatives can be predicted as follows:

2, 4-di-Cl-TFMS ($\Sigma\pi = +1.82$, $\Sigma\sigma = +0.454$)

Weed	$A(\Sigma\pi)^2$	$B(\Sigma\pi)$	$\rho(\Sigma\sigma)$	D	$\log(1/LD_{90})$	LD_{90} $\left(\frac{mole}{acre}\right)$	
Foxtail	−3.355	+0.604	+1.377	−1.079	−2.454	284.25	(32)
Wild Mustard	−5.942	+2.319	+1.594	−0.666	−2.695	495.81	(33)
Cheat Grass	−5.075	+1.390	+1.872	−0.818	−2.630	426.85	(34)

2,4-di-F-TFMS ($\Sigma\pi = +0.40$, $\Sigma\sigma = +0.124$)

Weed	$A(\Sigma\pi)^2$	$B(\Sigma\pi)$	$\rho(\Sigma\sigma)$	D	$\log(1/LD_{90})$	LD_{90} $\left(\frac{mole}{acre}\right)$	
Foxtail	−0.162	+0.132	+0.376	−1.079	−0.732	5.40	(35)
Wild Mustard	−0.287	+0.510	+0.435	−0.666	−0.008	1.02	(36)
Cheat Grass	−0.245	+0.306	+0.511	−0.818	−0.246	1.76	(37)

Comparing the above calculated LD_{90} values with the experimental data of Trepka *et al.* (*5*) reported in Table I, we note that our derived Hansch relationships satisfactorily predict the high grass and broadleaf

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

activity of the 2,4-di-F–TFMS derivative (very low LD_{90} concentrations). The grass and broadleaf activity of the 2,4-di-Cl–TFMS derivative is poorly predicted, however. It is evident from Equations 32–34 and Figures 2a, 3a, and 4a that the poorly predicted activity of 2,4-di-Cl–TFMS on all weed species is caused entirely by unfavorable partitioning characteristics. $\Sigma\pi(4\text{-Cl}) = +1.82$, for example, and this estimated value lies far from the optimum π_o values for all weed species (*see* Table XI).

It appeared reasonable, however, that the placement of an electronegative Cl group ortho to the acidic trifluoromethanesulfonamido parent group might well affect the partitioning characteristics of the overall molecule in ways that could not be accounted for by a simple partitioning additivity principle (*i.e.*, by assuming that π contributions for two 4-Cl substituents adequately represent the partitioning characteristics of a 2-Cl and 4-Cl substituent). As a check on the additivity of π for the 2,4-di-Cl– and 2,4-di-F–TFMS derivatives, the octanol/water partition coefficients of these compounds were experimentally determined using the procedure outlined in the Experimental section. The following results accurate to within ±2% were obtained.

Table XIV. Octanol/Water Partition Coefficients for Disubstituted TFMS Derivatives

TFMS Derivative	*log* P_X	$\Sigma\pi = log\ P_X - log\ P_H$
2,4-di-Cl–	3.88	+0.83
2,4-di-F–	3.06	+0.01
2-CH_3–, 4-Cl–	4.26	+1.21

The $\Sigma\pi$ values in Table XIV were calculated by subtracting log $P_H = 3.05$ for the original parent compound (in the absence of Tween 80) from the measured log P_X values for the disubstituted compounds. The $\Sigma\pi$ values tabulated above for the 2,4-di-F– and 2,4-di-Cl–TFMS derivatives thus represent the actual experimental effects on partitioning that 2,4-di-F or 2,4-di-Cl substitution have on the parent molecule. Comparing the above experimental $\Sigma\pi$ values with $\Sigma\pi(4\text{-Cl}) = +1.82$ and $\Sigma\pi(4\text{-F}) = +0.40$ used in the calculations of Equations 32–37, we note that the experimentally measured $\Sigma\pi$ is considerably lower than the estimate in both cases. For 2,4-di-Cl substitution, for example, the addition of the second Cl in the ortho position renders the molecule more hydrophilic (*i.e.*, $\pi(4\text{-Cl}) = +0.91$, $\Sigma\pi(2\text{-Cl} + 4\text{-Cl}) = +0.83$). It would appear reasonable that electronegative ortho-Cl substitution enhances ionization of the parent TFMS side chain, thus leading to a slight decrease in lipophilicity of the compound. A similar argument probably applies to the 2,4-di-F–TFMS derivative.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

If we continue to assume the additivity of σ but use the new experimentally measured $\Sigma\pi$ values in Table XIV for the 2,4-di-F– and 2,4-di-Cl–TFMS derivatives, a repeat of the calculations of Equations 32–37 yields the following results:

2,4-di-Cl–TFMS ($\Sigma\pi = +0.83$, $\Sigma\sigma = +0.454$)

Weed	$A(\Sigma\pi)^2$	$B(\Sigma\pi)$	$\rho(\Sigma\sigma)$	D	$\log(1/LD_{90})$	LD_{90} $\left(\frac{mole}{acre}\right)$	
Foxtail	−0.698	+0.276	+1.377	−1.079	−0.125	1.33	(38)
Wild Mustard	−1.236	+1.057	+1.594	−0.666	+0.750	0.18	(39)
Cheat Grass	−1.055	+0.634	+1.872	−0.818	+0.633	0.23	(40)

2,4-di-F–TFMS ($\Sigma\pi$ *+0.01*, $\Sigma\sigma = +0.124$)

Weed	$A(\Sigma\pi)^2$	$B(\Sigma\pi)$	$(\rho\Sigma\sigma)$	D	$\log(1/LD_{90})$	LD_{90} $\left(\frac{mole}{acre}\right)$	
Foxtail	−0.0001	+0.003	+0.376	−1.079	−0.700	5.01	(41)
Wild Mustard	−0.0002	+0.013	+0.435	−0.666	−0.218	1.65	(42)
Cheat Grass	−0.0002	+0.008	+0.511	−0.818	−0.299	1.99	(43)

It is evident from Equations 38–43 that use of the experimental $\Sigma\pi$ values from Table XIV leads to a correct prediction of the high grass and broadleaf activity of both the 2,4-di-Cl– and 2,4-di-F–TFMS derivatives. The use of the experimentally measured $\Sigma\pi$ for the 2,4-di-F derivative calculation did not lead to significant new changes in predicted activity because both $\Sigma\pi$ (experimental) and $\Sigma\pi$ (estimate) lie near the π_o values for the various weed types (*see* Table XI and Figures 2a, 3a, and 4a). The use of the measured $\Sigma\pi$ for the 2,4-di-Cl–TFMS derivative, however, leads to large predicted changes in activity because the experimental $\Sigma\pi$ is now much closer to the optimum π_o values for the various weed types than is $\Sigma\pi$ (estimate). Also, in the 2,4-di-Cl case, $\Sigma\sigma = +0.454$ is quite favorable and leads to a large, positive $\rho\Sigma\sigma$ term in both the grass and broadleaf Hansch relationships of Equations 38–40. The high grass and broadleaf activity of the 2,4-di-Cl derivative thus appears to be caused primarily by the large, positive $\Sigma\sigma$ and secondarily by reasonably good partitioning characteristics. The high activity of the 2,4-di-F derivative appears to be attributable to good partitioning characteristics and to a lesser extent to the positive contributions to $\log(1/LD_{90})$ from $\rho\Sigma\sigma$ (*cf.* positions of $\Sigma\pi$ and $\Sigma\sigma$ on Figures 2a, 3a, and 4a for the 2,4-di-F– and 2,4-di-Cl–TFMS derivatives).

The above treatment demonstrates the utility of the Hansch relationships of Table XI in predicting the grass and broadleaf activity of several disubstituted TFMS compounds. It also demonstrates the dangers in-

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

herent in assuming the additivity of π (at least for certain electronegative groups substituted ortho to acidic parent side chains). This is not meant to imply that the additivity of π does not hold in other cases, however. We have, for example, experimentally measured log P for the $2CH_3$–, 4Cl–TFMS derivative and found the measured $\Sigma\pi = +1.21$ for this disubstituted derivative (*see* Table XIV). This value, although again lower, is not too far removed from the estimate $\Sigma\pi = \pi(4\text{-Cl}) + \pi(4\text{-}CH_3) = +0.91 + 0.55 = +1.46$ calculated by an additivity relationship. Caution in applying additivity relationships is recommended; where possible, π and $\Sigma\pi$ values should be measured experimentally.

Implications of the TFMS Correlations. META-TFMS ACTIVITY ON WILD MUSTARD IN THE PRESENCE OF TWEEN 80. Table XI and Figure 3d showed that the activity of the meta-TFMS herbicides on Wild Mustard in the presence of Tween 80 was anomalous (relative to the other weed types and formulation conditions examined) in the following ways:

(a) The meta-TFMS derivatives exhibited a minimum in herbicidal activity in the π range of our study. For all other correlation conditions investigated, a maximum in activity was obtained over this same experimental π range (*cf.* Figures 2–4 and π_o values in Table XI).

(b) Inclusion of the H-substituted parent compound in the data pool for this particular correlation resulted in only an 8% loss in percent correlation (r^2)—*see* Table IX. In every other case, inclusion of the parent TFMS compound with 3-substituted TFMS derivatives in the stepwise regression procedure resulted in percent correlation decreases of from 25 to 40%.

(c) Inclusion of the 3-SCH_3–TFMS derivative in the data pool for the above-cited Wild Mustard correlation resulted in a decrease in percent correlation of less than 5% compared with typical decreases of 20–50% in all but one other case upon inclusion of the methylthio derivative (*see* Results).

(d) For the above case only, no dependence of the final Hansch correlation equation on the Hammett σ constant was observed (*cf.* equations in Table XI and *see* Figure 3d).

Item d implies that in terms of the Hansch model of Equation 1, partitioning STEP 1 is the primary rate-controlling process characterizing the herbicidal action of the 3-TFMS compounds on Wild Mustard in the presence of Tween 80. Since all the other Hansch relationships in Table XI include fairly significant $\rho\sigma$ contributions, partitioning as well as other rate processes (possibly more intimately connected with the receptor site within the plant or seed) must be involved in determining overall herbicidal activity in these latter cases. One may speculate that the anomalous observations (a–c) above are the direct consequence of (d)—the lack of Hammett σ dependence. If the $\rho\sigma$ term in the Hansch equation does indeed reflect rate or equilibrium events occurring at or near the herbicidal site of action within the plant or seed (as is often assumed but not

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

necessarily proved for the Hansch model), then items a–c are connected with events taking place in the vicinity of the herbicidal receptor site. For example, partitioning as well as rate processes occurring at or near the receptor site may contribute to determining the position and broadness of the π optimum for the activity surfaces in all other cases but the one cited here. In like manner, one reason that the H-substituted parent compound fits in more satisfactorily as a member of the 4-substituted TFMS family than with the 3-TFMS compounds (*see* Table IX and Table X) may be caused by a similarity in the mode of binding (or reaction) of the parent compound at a receptor site (possibly an enzymic site) within the plant or seed in which the parent TFMS herbicide more closely resembles the para-substituted than the meta-substituted series members. Alternatively, the parent compound may be metabolically converted to a 4-substituted TFMS derivative following absorption by the plant or seed but prior to reaction at the receptor site. This suggests that the 3- and 4-substituted TFMS compounds either bind to a given site with different orientations or react with it at different rates. On the other hand, the meta- and para-substituted derivatives may interact with different receptor sites within the plant or seed. If so, this could be the reason for the observed herbicidal activity differences between 3- and 4-substituted TFMS compounds. Possibly the inclusion of a steric term (E_S) or an additional electronic parameter in the Hansch equation for each substituent would account for the activity differences between the two herbicidal families so that a single correlation could be used to explain combined 3- and 4-TFMS activity against a given weed type. An approach of this type has not been undertaken with the TFMS herbicides but is planned.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

HERBICIDAL ACTIVITY OF METHYLTHIO–TFMS DERIVATIVES. Except for meta- and para-TFMS compounds acting on Wild Mustard in the presence of Tween 80 (the activity of both series of which depends strongly on partitioning parameters—*see* Figures 3c–3d), the methylthio–TFMS herbicidal data points could not be satisfactorily correlated by any of the final Hansch relationships in Table XI. This implies that the source of the anomalous 4-SCH_3–TFMS and 3-SCH_3–TFMS activity lies at (or in the vicinity of) the receptor site of herbicidal action within the plant or seed rather than at a partitioning site. In terms of the Hansch model of Equation 1, this suggests that the methylthio-TFMS derivatives may possibly function herbicidally *via* a mode of action different from the other compounds examined (assuming, of course, that the $-SCH_3$ grouping remains chemically and metabolically intact at least until it reaches the receptor site).

An alternative (and probably more likely) explanation for the anomalous herbicidal activity of the methylthio–TFMS derivatives exists [sug-

gested by K. H. Büchel and J. A. Durden]. Fukuto and co-workers (*31*, *32*), in a series of structure–activity correlation studies, observed anomalously high activity for 4-SCH_3-substituted diethylphenyl phosphate anticholinesterase insecticides. They attributed this unexpectedly high activity to the *in vivo* metabolic oxidation of the methylthio moiety to the strongly electron-withdrawing $-SOCH_3$ ($\sigma^- = +0.567$) and/or $-SO_2CH_3$ ($\sigma^- = +1.049$) groups, both of these latter derivatives being highly toxic to insects. Examination of our methylthio structure–activity data strongly suggests that a similar *in vivo* oxidation may be occurring with the SCH_3–TFMS compounds examined in our study. Table XV lists experimentally measured and calculated LD_{90} values for the meta- and para-substituted methylthio–TFMS derivatives for each of the weed types examined in the presence and absence of Tween 80. The calculated values were determined using the Hansch relationships of Table XI, from which herbicidal activity data for the methylthio–TFMS derivatives were omitted when the equations were computer-derived. For comparison, experimentally determined and calculated LD_{90} values (determined by using the appropriate relationships in Table XI) are included for both the 3- and 4-substituted methylsulfonyl–TFMS compounds. As seen in Table XV, the experimental LD_{90} values for the 4-SO_2CH_3–TFMS and

Table XV. Calculated and Experimental LD_{90} Values for Methylthio– and Methylsulfonyl–TFMS Derivatives

			Methylthio (SCH_3)		*Methylsulfonyl* (SO_2CH_3)	
Weed Type	*0.1% Tween 80 Present (?)*	*TFMS Series*	LD_{90} (*Calc.*) [a] *mole/acre*	LD_{90} (*Exp.*) [b] *mole/acre*	LD_{90} (*Calc.*) [a] *mole/acre*	LD_{90} (*Exp.*) [b] *mole/acre*
Foxtail Grass	No	Para	29.9	5.1	2.30	2.42
	No	Meta	38.3	<2.1	1.56	1.50
	Yes	Para	34.6	5.1	2.53	2.64
	Yes	Meta	57.9	2.1	1.16	1.17
Cheat Grass	No	Para	16.4	2.6	2.22	2.21
	No	Meta	55.7	<2.1	7.80	<1.90
Wild Mustard	No	Para	6.4	42.0	29.9	31.6
	No	Meta	22.2	9.5	19.2	17.4
	Yes	Para	18.0	25.2	83.2	79.0
	Yes	Meta	19.9	17.2	28.4	26.8

[a] Calculated LD_{90} values are obtained by substituting appropriate π (or π') and σ values for the $-SCH_3$ and $-SO_2CH_3$ derivatives from Table IV and VI into the corresponding final Hansch relationships for each weed type and set of surfactant conditions in Table XI. 3-SCH_3–TFMS and 4-SCH_3–TFMS data points were omitted in deriving the equations of Table XI.

[b] Experimental LD_{90} values are those listed in Table VII for the methylthio–TFMS and methylsulfonyl–TFMS derivatives.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

3-SO_2CH_3–TFMS derivatives are, for the most part, well predicted for all weed types in the presence and absence of surfactant. The same cannot be said for the 4-SCH_3–TFMS and 3-SCH_3–TFMS series members. For both Foxtail and Cheat Grass, the LD_{90}'s calculated by the appropriate equations in Table XI range from 5–25 times too high. The Wild Mustard LD_{90} values for the non-surfactant case are also poorly predicted. Referring to Table XV, comparison of the experimentally measured LD_{90} values for the methylthio–TFMS derivatives acting on Foxtail and Cheat Grass with the calculated and/or experimental LD_{90} values for the methylsulfonyl (–SO_2CH_3) derivatives reveals that for the most part they are very similar. As suggested by the work of Fukuto *et al.* (*31*, *32*), it seems reasonable to conclude that for both Foxtail (in the presence and absence of Tween 80) and Cheat Grass, the methylthio side chain is converted *in vivo* (*see below*) to methylsulfonyl (–SO_2CH_3) or possibly methylsulfinyl (–$SOCH_3$). Since both of these side chains are characterized by considerably more positive Hammett sigma constants than methylthio [σ (4SCH_3) = −0.047; σ (3SCH_3) = +0.144; σ (4SO_2CH_3) = +0.728; σ (3SO_2CH_3) = +0.647; σ (4$SOCH_3$) = +0.567; σ (3$SOCH_3$) = +0.551], one would expect metabolic conversion of methylthio to –$SOCH_3$ or –SO_2CH_3 prior to ultimate reaction at the receptor site to be accompanied by significant increases in herbicidal activity. (Note in Table XI, for example, that for both Foxtail and Cheat Grass the σ coefficient, ρ, is positive for all sets of experimental conditions examined so that positive increases in σ result in enhanced activity—*i.e.*, larger $\log(1/LD_{90})$ values).

Referring to the broadleaf data in Table XV for Wild Mustard in the absence of surfactant, experimental LD_{90} values for the methylthio–TFMS compounds are again poorly predicted by the Hansch relationships of Table XI. As with Foxtail and Cheat Grass, the experimental LD_{90} values for 3-SCH_3–TFMS and 4-SCH_3–TFMS lie much closer to corresponding methylsulfonyl-TFMS LD_{90}'s than they do to the calculated methylthio–TFMS LD_{90} values. Thus, it appears that oxidation of TFMS methylthio (–SCH_3) side chains to methylsulfinyl (–$SOCH_3$) or methylsulfonyl (–SO_2CH_3) before reaction at the receptor site is highly probable for the non-surfactant Wild Mustard cases.

The only two cases in Table XV in which the experimental methylthio LD_{90} values are satisfactorily predicted by the Hansch relationships of Table XI are for 3-SCH_3–TFMS and 4-SCH_3–TFMS acting on Wild Mustard in the presence of Tween 80. Examination of the appropriate equations in Table XI and of the plots in Figures 3c and 3d for Wild Mustard reveals that for the 4-SCH_3–TFMS surfactant case, herbicidal activity depends strongly on partitioning events (note the large magnitude of the π^2 and π coefficients) while for the 3-SCH_3–TFMS derivative

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

in the presence of Tween 80, partitioning processes completely govern the observed herbicidal activity (*i.e.*, $\log(1/LD_{90})$ is independent of σ). Since the Wild Mustard activity of all TFMS derivatives is dominated by apparently rate-limiting partitioning processes in the presence of Tween 80, any methylthio oxidative conversion to $-SOCH_3$ or $-SO_2CH_3$ which may occur after the TFMS herbicide has been partitioned from an extracellular to an intracellular phase would probably not appear experimentally as a deviation from the expected behavior.

Comparison of the Foxtail and Wild Mustard herbicidal activity data for the TFMS compounds in Table XV gathered in the presence of Tween 80 strongly suggests that if oxidative conversion of the $-SCH_3$ side chain to $-SOCH_3$ or $-SO_2CH_3$ does occur, it must do so intracellularly rather than extracellularly (*i.e.*, outside the plant or seed). Since the Foxtail and Wild Mustard herbicidal evaluations in the presence of surfactant were carried out simultaneously in the same plastic flower pot (*see* Experimental), extracellular oxidation of methylthio to methylsulfonyl or methylsulfinyl would have occurred equally for both weed types. This should have resulted in the experimental LD_{90}'s for 4-SCH_3–TFMS and 3-SCH_3–TFMS acting on Wild Mustard in the presence of surfactant coinciding or nearly coinciding with corresponding LD_{90} values for the 4-SO_2CH_3–TFMS and 3-SO_2CH_3–TFMS derivatives as they do in the Foxtail surfactant case (*see* Table XV). Although methylthio–TFMS derivatives might possibly be protected from *in vivo* oxidation by Tween 80 absorbed by the Wild Mustard seeds (micelle formation), a process of this type would not be observable experimentally since slower rate-determining partitioning events control the overall herbicidal activity as previously explained.

Thus, it seems that the anomalous methylthio–TFMS herbicidal activity we have observed in our structure–activity studies on grass and broadleaf weeds is attributable to *in vivo* conversion of the $-SCH_3$ side chain to more highly oxidized forms ($-SOCH_3$, $-SO_2CH_3$). Alternatively, methylthio–TFMS derivatives may function herbicidally *via* a completely different mode of action than the other TFMS derivatives examined. In light of the arguments presented, however, this seems to be a much less likely possibility than the *in vivo* oxidation proposal.

σ^- VALUES. Hammett (*12*) and Jaffe (*13*) have recommended the use of special σ^- values (sometimes denoted σ_{para}) for certain strongly electron-attracting groups (like 4-SO_2CH_3) attached at the *para* position of a benzene ring containing a strongly electron-donating parent group like $-NH_2$ or $-OH$. The use of σ^- (instead of regular σ) values for these specific para-substituents in aniline and phenol is thus usually advised. [σ^- values have been tabulated and are available (*13*).] Since the parent TFMS compound (I) is a specially substituted aniline, it was not clear

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

whether σ^- or regular Hammett σ values should be used in the Hansch correlations of our study. Since the strongly electron-attracting trifluoromethanesulfonyl group was attached directly to the aniline amino group ($-NH-SO_2CF_3$) of our parent compound, we suspected that the ability of the nitrogen to donate its lone-pair electrons to the aromatic system would be severely impaired (in which case the σ^- values would not apply). Accordingly, all the correlations undertaken in this study were carried out using both σ and the special σ^- values (where applicable). For the series of TFMS compounds evaluated in our work, only the $4\text{-}SO_2CH_3$–, 4F–, and $4\text{-}OCH_3$–TFMS derivatives and the correlations in which they were involved were applicable. Although the results are not tabulated here, in every case better fits (as judged by the statistical parameters r, r^2, SE, and F) were obtained when the normal Hammett σ (rather than σ^-) was used to correlate the data. Hence, the normal Hammett σ constants in Table IV were used to correlate all the data discussed here. The fact that the TFMS herbicidal data are better correlated by σ than σ^- reflects the strong electron-attracting power of the $-SO_2CF_3$ group, which, on the basis of our analysis, appears to inhibit the donation of the nitrogen lone-pair electrons to the aromatic system, as initially suspected. The $-SO_2CF_3$ group is, of course, also responsible for the acidity ($pK_a = 4.45$) of the parent TFMS compound and its derivatives. The source of the activity differences between the 3- and 4-substituted TFMS herbicides may lie in the direct resonance effects which 4-substituents have on the acidity of the parent $-NHSO_2CF_3$ grouping. Meta substituents would be expected to exert a lesser inductive effect on the acidity of the parent side chain. If this were the case, inclusion of a term involving the pK_a of each TFMS compound in the Hansch equation might help account for activity differences between the meta and para series.

CORRELATIONS USING LD_{50} VALUES. All computer correlations discussed here were carried out using herbicidal activity data expressed as LD_{90} values. From the Log-Probit plots of our activity data (*see* Results) we were also able to determine LD_{50} values for all TFMS derivatives under all experimental conditions examined. For comparative purposes, all computer correlations discussed herein for LD_{90} values were also undertaken using activity data expressed in terms of LD_{50} values. Except for obvious variations in the π and σ parameter coefficients in the Hansch equation necessary to account for the fact that the LD_{50} values for all derivatives will be less than their corresponding LD_{90} values, the same overall conclusions were predicted whether LD_{50} or LD_{90} values were used in carrying out the structure–activity correlations. For example, separation of TFMS derivatives into meta and para series, anomalous methylthio–TFMS behavior, H-substituted parent compound placement with the para-TFMS derivatives, π_o values, etc., were all equally well

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

predicted using LD_{50} values instead of LD_{90} values in fitting the Hansch correlation equation. This good, expected agreement between conclusions and predictions using LD_{50} or LD_{90} expressions of herbicidal activity substantiates the validity of our results.

EFFECT OF TWEEN 80 ON THE RELATIVE HERBICIDAL POTENCY OF TFMS COMPOUNDS. Figure 6 illustrates the effect of Tween 80 on the π values of meta- and para-TFMS series members. All π values are referenced against the parent compound in the absence of surfactant (*cf.* Table VI). The scale at the top of Figure 6 lists π values in order of increasing lipophilicity for the para substituents. A corresponding scale for meta substituents is at the bottom of the figure. The hydrophilic end of each scale is at the left on both scales, and the hydrophobic or lipophilic end is at the right. The top portion of each scale corresponds to π values measured in the absence of surfactant while the lower portion of each scale corresponds to π' values determined in the presence of 0.1% Tween 80. Several points in Figure 6 are worth noting:

(1) In the absence of Tween 80 both the para- and meta-TFMS π values are evenly distributed along the uppermost π axes. When Tween 80 is added, the π scales for both the para and meta derivatives are severely compressed, as shown in the lower portion of each scale.

(2) For both meta and para derivatives, all π values are shifted to more negative values when surfactant is added, indicating that the derivatives become more water soluble.

(3) On both the para and meta scales, the π values of the lipophilic substituents like $-CF_3$ are shifted much more drastically by surfactant addition than are the π values of the hydrophilic substituents like $-SO_2CH_3$.

(4) The para-TFMS π values are more compressed by surfactant addition than are the meta-TFMS π values.

Of considerable interest also is the effect that surfactant has on the position of the various π optima for the Foxtail and Wild Mustard activity surfaces (*see* Figures 2 and 3). Table XI shows that π_o for the para-TFMS derivatives acting on Foxtail in the absence of surfactant is $\pi_o = +0.16$, while in the presence of Tween 80 it is shifted to $\pi_o' = -0.77$. Referring now to Figure 6, note that the π_o value for the para substituents in the non-surfactant case falls between the π values for the 4-F– and 4-OCH_3–TFMS substituents. In the presence of surfactant, the π optimum is shifted toward more hydrophilic substituents and is now located between the π values of the 4-OCH_3– and 4-SO_2CH_3– substituents.

Similarly, for Wild Mustard the π optimum from Table XI for the para-TFMS derivatives in the absence of Tween 80 is $\pi_o = +0.36$ while in its presence the value is shifted to -0.59. Referring again to Figure 6, note that the π optimum is located between the 4-CH_3– and 4-F–TFMS derivatives in the absence of surfactant. In the presence of Tween 80

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

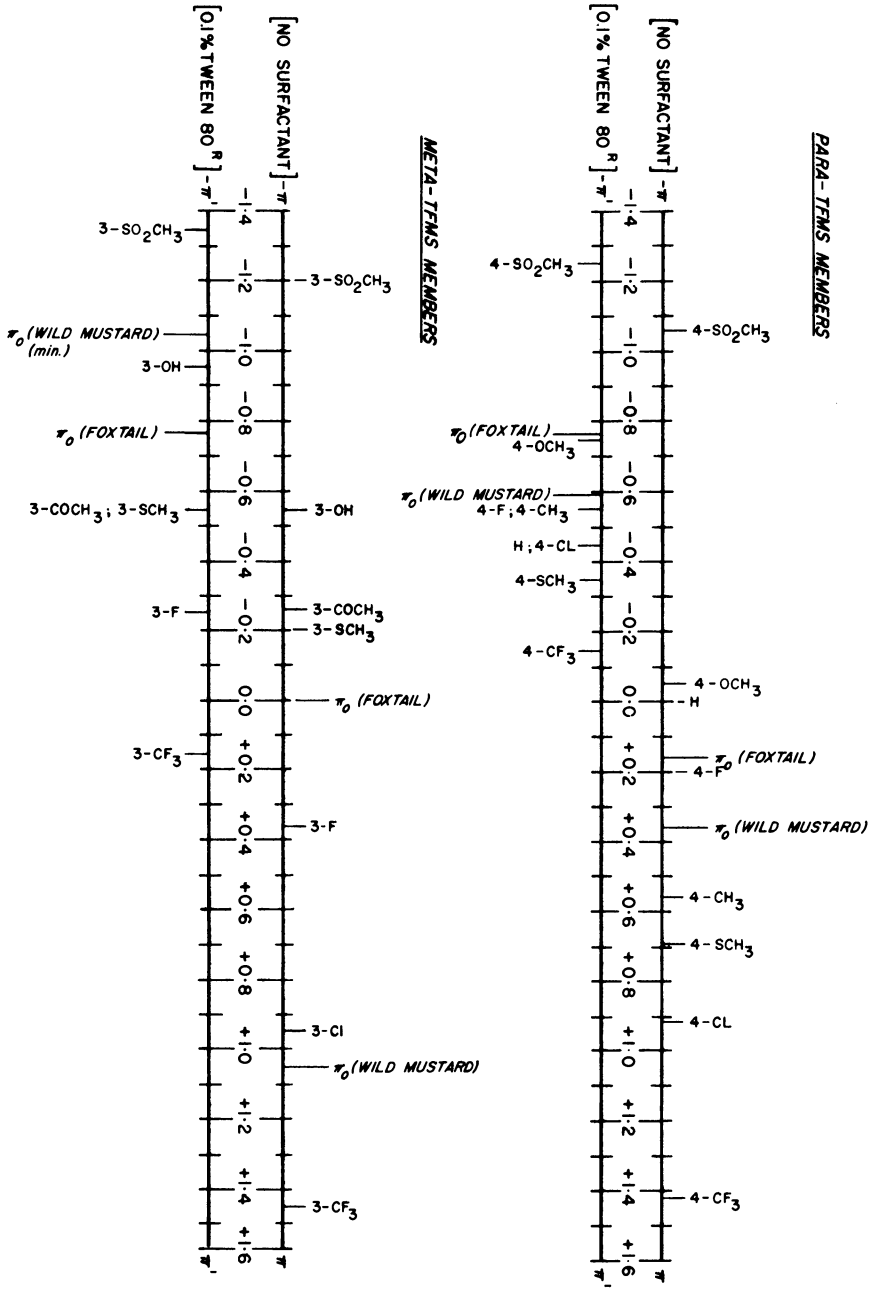


Figure 6. π values for para- and meta-TFMS herbicides obtained in the absence (π) and presence (π′) of 0.1% Tween 80 (see Table VI for the definition of π and π′). π_o values are optimum π values corresponding to those shown in the activity surfaces of Figures 2 and 3 and in Table XI.

the π optimum is again shifted hydrophilically and now lies between the π values of the 4-F– and 4-OCH_3–TFMS series members.

The meta-TFMS derivatives acting on Foxtail and Wild Mustard exhibit similar hydrophilic shifts of π optima (Figure 6). For the meta derivatives acting on Foxtail, these shifts are less important on activity because of the smaller π dependence of the meta substituents relative to the para derivatives in the presence and absence of surfactant (*see* Figure 2). For Wild Mustard, however, the hydrophilic shifts in meta substituent π values are quite significant because of the large influence of Tween 80 on meta-TFMS partitioning characteristics (*see* Figure 3d).

The implications of these observations in herbicidal screening tests are important. A small amount of surfactant added to a herbicidal formulation to improve penetration of the herbicide into plant cells might do just that—*i.e.*, enhance the herbicidal activity. As the plots of Figures 2 and 3 show, however, it is also possible that herbicidal activity will be inhibited. The addition of surfactant in even small amounts could shift the optimum of activity in a given herbicidal family of compounds so that a potent compound in the absence of surfactant becomes ineffective in its activity towards a particular grass or broadleaf in the presence of surfactant. If a surfactant is routinely used to prepare all herbicidal formulations in screening tests (*i.e.*, evaluations carried out at a single relatively high constant application dosage—*e.g.*, 20 lb/acre), a relatively potent compound (in the absence of surfactant) might be classified as an inactive herbicide (in the presence of surfactant) and eliminated from further testing. Of course, the opposite effect is also possible—*i.e.*, a relatively poor herbicide in the absence of surfactant could become a potent one in its presence.

Proper comparison of the activity surfaces in Figures 2 and 3 for the meta–TFMS derivatives acting on Foxtail and Wild Mustard in the presence and absence of surfactant provides clues as to how a surfactant like Tween 80 can be used to improve their grass and broadleaf selectivity. For example, when surfactant is added to meta–TFMS formulations applied to Foxtail Grass, the right, lipophilic portion of the activity surface is bent down, and the grass activity of the more hydrophobic meta derivatives is strongly depressed (*cf.* Figures 2b and 2d). If Tween 80 is added to meta-TFMS formulations applied to Wild Mustard, however, the broadleaf activity of the more hydrophobic meta derivatives is greatly enhanced (*cf.* Figures 3b and 3d). From a selectivity standpoint, addition of surfactant thus can improve the pre-emergence activity of hydrophobic series members against broadleaf weeds like Wild Mustard (*see* Figure 3d) while simultaneously depressing grass activity (*see* Figure 2d). At the other extreme of the π scale, surfactant addition will enhance the grass activity of hydrophilic series members and depress

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

broadleaf activity, especially if a derivative with a large, positive σ constant is chosen (*e.g.*, 3-SO_2CH_3–TFMS).

Since our analysis has demonstrated that surfactant addition to herbicidal formulations can produce a variety of activity-inhibiting and activity-enhancing effects, it is prudent to avoid the indiscriminate use of surfactants as penetrant aids and formulating agents in herbicidal screening tests. Herbicidal evaluations in the presence and absence of surfactant coupled with regression analyses of the type discussed here should provide valuable clues as to the most advantageous use of surfactants in enhancing herbicidal activity and in providing better control of grass and broadleaf selectivity.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Summary

Hansch analyses *via* computer regression techniques were used to correlate the grass and broadleaf pre-emergence activity of a series of 15 meta- and para-substituted trifluoromethanesulfonanilides (TFMS) with structural changes in the parent molecule. The herbicidal activity of the 3-substituted series members differs from that of the 4-substituted derivatives. Each series satisfies a different form of the Hansch equation in its action on a particular weed type. The final correlation equations for Cheat Grass, Foxtail, and Wild Mustard differ. Subtle differences in the activity of the TFMS compounds against the two grass species Foxtail and Cheat Grass were clarified by using Hansch stepwise regression analyses.

Hansch analyses have shown that the surfactant Tween 80, when used in herbicidal formulations of TFMS derivatives at the 0.1% (w/v) concentration level, can produce one or more of the following effects in pre-emergence herbicidal tests, depending on the weed type and/or TFMS derivative under evaluation: (a) no effect, (b) enhancement of herbicidal activity, (c) inhibition of herbicidal activity, or (d) shifting of optimum herbicidal activity from one TFMS derivative to another within each series. Mathematical equations relating structure to activity have been derived for the TFMS compounds both for the surfactant present and surfactant absent cases.

Partitioning (π) effects are in general much more significant in controlling the herbicidal activity of 4-substituted TFMS compounds than they are for 3-substituted derivatives, especially in the absence of surfactant.

In general, substituents characterized by large, positive Hammett σ constants (electron-attracting substituents) exhibit enhanced pre-emergence herbicidal activity on all weed types tested (Foxtail, Cheat Grass,

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

Wild Mustard) relative to TFMS substituents characterized by small or negative σ constants.

Methylthio–TFMS derivatives (3-SCH_3–TFMS, 4-SCH_3–TFMS) exhibit anomalous behavior and may be oxidized *in vivo* to methylsulfinyl (–$SOCH_3$) or methylsulfonyl (–SO_2CH_3) before reacting at the receptor site within the plant or seed. Alternatively, they may function by a mode of action different from the other TFMS pre-emergence herbicides tested. The parent TFMS (H-substituted) appears to function as a 4-substituted series member rather than as a 3-substituted derivative. The predictive capabilities of the derived Hansch relationships were demonstrated for several specific cases.

Three-dimensional perspective plots which show the relationships between π, σ, and $\log(1/LD_{90})$ were prepared for all weed types examined. Herbicidal activity differences induced by surfactant addition to TFMS formulations are demonstrated by these plots.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch012

Acknowledgments

The author is indebted to Jon Belisle for carrying out the partition coefficient measurements and for stimulating discussions on various aspects of this subject. Thanks are also due J. Waddington and D. Pauly for carrying out the pre-emergence herbicidal tests and to J. K. Harrington and R. D. Trepka for providing samples of the TFMS herbicides used in this study. We are also grateful to Earl Cook and R. Borden for their advice on various mathematical aspects of this investigation. Thanks also to J. W. Van Valkenburg for many helpful discussions on problems associated with herbicidal formulations and to K. H. Büchel and J. Durden for several helpful suggestions.

Literature Cited

1. Brice, T. J., Trott, P. W., U.S. Patent **2,732,398** (Jan. 24, 1956).
2. Burdon, J., Farazmand, I., Stacy, M., Tatlow, J. C., *J. Chem. Soc.* (1957) 2574.
3. Farrar, W. V., *J. Chem. Soc.* (1960) 3058.
4. Harrington, J. K., Robertson, J. E., Kvam, D. C., Hamilton, R. R., McGurran, K. T., Trancik, R. J., Swingle, K. F., Moore, G. G. I., Gerster, J. F., *J. Med. Chem.* (1970) **13**, 137.
5. Trepka, R., Harrington, J. K., Robertson, J. E., Waddington, J. T., *J. Agr. Food Chem.* (1970) **18**, 1176.
6. Hansch, C., "Annual Reports in Medicinal Chemistry," C. K. Cain, Ed., Vol. 34, p. 347, Academic, New York, 1966.
7. *Ibid.*, Vol. 35, p. 348, 1967.
8. Hansch, C., *Proc. Intern. Pharmacol. Meetg., 3rd* (1967) **7**, 141.
9. Hansch, C., *Farmaco, Sci. Ed.* (1968) **23**, 293.
10. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **86**, 1616.
11. Fujita, T., Iwasa, J., Hansch, C., *J. Amer. Chem. Soc.* (1964) **86**, 5175.

12. Hammett, L. P., "Physical-Organic Chemistry," p. 184, McGraw-Hill, New York, 1940.
13. Jaffe, H. H., *Chem. Rev.* (1953) **53,** 191.
14. Hansch, C., Kiehs, K., Lawrence, G. L., *J. Amer. Chem. Soc.* (1965) **87,** 5770.
15. Hansch, C., Deutsch, E. W., Smith, R. N., *J. Amer. Chem. Soc.* (1965) **87,** 2738.
16. Kiehs, K., Hansch, C., Moore, L., *Biochemistry* (1966) **5,** 2602.
17. Helmer, F., Kiehs, K., Hansch C., *Biochemistry* (1969) **7,** 2858.
18. Hansch, C., Stewart, A. R., Anderson, S. M., Bentley, D., *J. Med. Chem.* (1968) **11,** 1.
19. Lien, E., Hansch, C., Anderson, S. M., *J. Med. Chem.* (1968) **11,** 430.
20. Hansch, C., Lien, E., *Biochem. Pharmacol.* (1968) **17,** 709.
21. Neely, W. B., Whitney, W. K., *J. Agr. Food Chem.* (1968) **16,** 571.
22. Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F., Streich, M., *J. Amer. Chem. Soc.* (1963) **85,** 2817.
23. Cammarata, A., Yau, S. J., *J. Med. Chem.* (1970) **13,** 93.
24. Martin, Y. C., *J. Med. Chem.* (1970) **13,** 145.
25. Purcell, W. P., Beasley, J. G., Quintana, R. P., Singer, J. A., *J. Med. Chem.* (1966) **9,** 297.
26. Hansch, C., Deutsch, E., *Biochem. Biophys. Acta* (1966) **126,** 117.
27. Goldstein, A., Aronow, L., Kalman, S., "Principles of Drug Action," Chap. 5, Harper and Row, New York, 1968.
28. Trevan, T. W., *Proc. Roy. Soc. B.* (1927) **101,** 483.
29. Finney, D. J., "Probit Analysis," 2nd ed., Cambridge Univ. Press, London, 1952.
30. Fisher, R. A., Yates, F., "Statistical Tables for Biological, Agricultural and Medical Research," Oliver and Boyd, London, 1938.
31. Fukuto, T. R., Metcalf, R. L., *J. Agr. Food Chem.* (1956) **4,** 930.
32. Jones, R. L., Metcalf, R. L., Fukuto, T. R., *J. Econ. Entomol.* (1969) **62,** 801.
33. Leo, A., Hansch, C., Elkins, D., *Chem. Rev.* (1971) **71,** 525.
34. Ostle, B., "Statistics in Research," pp. 529–543, Iowa State University Press, Ames, 1963.

RECEIVED July 27, 1971.

13

The Effect of a Penetrant Aid on Pre-Emergence Herbicidal Activity of Trifluoromethanesulfonanilides

WADE VAN VALKENBURG and ANTHONY F. YAPEL, JR.

3M Company, St. Paul Minnesota 55133

Meta and para monosubstituted trifluoromethanesulfonanilides were tested for pre-emergence herbicidal activity against Foxtail and Wild Mustard. Activity was determined in the presence and absence of 0.1% by weight of a surfactant penetrant aid (Tween 80). The surfactant enhanced biological activity in some series members and reduced it in others. The change in biological activity caused by the surfactant was related primarily to the change in octanol/water partition coefficient induced by the surfactant. For para-substituted series members this change in partition coefficient was strongly correlated with variations in the Hammett sigma constant.

Surface-active agents (also known as surfactants, wetting agents, adjuvants, and spreader-stickers) are capable of modifying the observed biological activity of a herbicide. Jansen *et al.* (*1*) tested the foliar activity of three herbicides in combination with 63 different surfactants. Some of these surfactants enhanced biological activity while others decreased it.

Becher and Becher (*2*) measured the spreading pressure, π, of a series of surfactants on plant and synthetic surfaces. The surfactants were characterized by HLB (hydrophile-lipophile balance) values (*3*) similar to those of Jansen's adjuvants. The relationship between the spreading pressure and the contact angle can be expressed as $\pi = \gamma_L \cos \theta$, where γ_L is the surface tension of the liquid and θ is the contact angle between the liquid and a solid surface. Becher and Becher demonstrated that surfactants in their series which exhibited maximum spreading pres-

sure corresponded quite closely to those of Jansen's adjuvants of similar HLB which optimized herbicidal activity. Indeed, plots of either spreading pressure or Jansen's activity index *vs.* HLB yielded curves of similar shape.

Foy and Smith (*4*) recently reviewed the role of surfactants in modifying the activity of herbicidal sprays (66 references). These authors studied the adjuvant effects of a surfactant based on nonyl phenol with varying amounts of ethylene oxide appended. Activity was enhanced at an optimum ethylene oxide chain length. This is another way of stating that there is an optimum surfactant HLB for enhancing herbicidal activity on a given species of weed.

The hydrophile-lipophile balance is related to the solubility of a surfactant. At high HLB the surfactant is very water soluble while at low HLB the surfactant is very lipid soluble. Because HLB is solubility related, it is in turn related to the partition coefficient (or ratio of the solubility of the surfactant in a lipid phase to its solubility in an aqueous phase). A high HLB value suggests a low oil/water partition coefficient, and conversely a low HLB shows a high partition coefficient. Hence, an optimum surfactant HLB for enhancing biological activity also implies an optimum partition coefficient for activity enhancement.

Hansch and Fujita (*5*) have proposed a model for biological activity which may be expressed by the following equation:

$$\log \frac{1}{C} = a\pi^2 + b\pi + \rho\sigma + d \qquad (1)$$

where π = log P-log P°

P = octanol/water partition coefficient of a biologically active compound

P° = partition coefficient of a reference compound

σ = Hammett sigma constant

C = a concentration to yield a given biological effect, often expressed as LD_{50} or LD_{90}

a, b, ρ, and d are fitting parameters.

The Hansch equation mathematically describes two events controlling the biological activity of a compound. The first involves cell or membrane penetration (influenced by the partition coefficient), and the second involves interaction of the active molecule at a receptor site (modified by variations in the Hammett sigma constant).

Concerning Equation 1, the principal effect of adding a wetting agent to a herbicidal solution is to change the partition coefficient of the

herbicide from P to P'. Thus the biological activity of a herbicidal solution containing a surfactant may be represented as

$$\log\left(\frac{1}{LD_{90}'}\right) = a'(\log P' - \log P^{\circ})^2 + b'(\log P' - \log P^{\circ}) + \rho'\sigma + d' \qquad (2)$$

For this system the partition coefficient of the parent compound in the absence of surfactant is chosen as the reference state. Herbicidal activity is expressed as LD_{90} (or the concentration in moles/acre necessary to control 90% of the species).

If Equation 2 is subtracted from Equation 1, a complex equation with many nonfactorable terms results. For simplification, let us assume that the phenomenological form of Equations 1 and 2 can be used empirically to characterize the change in LD_{90} resulting from the addition of a surfactant to a herbicidal system. This empirical relationship can be expressed as

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = A(\log P - \log P')^2 + B(\log P - \log P') + C\sigma + D$$

If we let $\chi = (\log P - \log P')$ the above equation becomes

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = A^2\chi + B\chi + C\sigma + D \qquad (3)$$

Equation 3 thus represents the change in biological activity of a given compound caused by the presence of a surface-active agent. If the right side of the equation is negative, activity is enhanced. If it is positive, activity is reduced.

Materials and Methods

The details of the experimental procedure have been given in the preceding chapter (*6*) and are only summarized here.

Test Plants. Seeds of the grass Giant Foxtail *(Setaria sp.)* and the broadleaf Wild Mustard *(Brassica kaber)* were planted in a mixed soil (*7*) and bottom-watered until emergence occurred. Top watering was then used for the rest of the test.

Pre-Emergence Studies. Pre-emergence herbicidal evaluations of the 15 trifluoromethanesulfonanilide (TFMS) derivatives listed in Table I were conducted in an artificially illuminated greenhouse. Each of the candidate compounds was applied to appropriately seeded soil samples as an aqueous drench at three or four dosage levels ranging from 1.25 to 20 lb/acre. All drenches contained 1% acetone (w/v) to aid dispersion of the herbicide in water. Emergence and plant vigor were measured after a 21-day growing period. Herbicidal dosages were converted to a mole/acre concentration designation for purposes of Hansch correlation.

LD_{50} and LD_{90} values were obtained for each herbicide from log-probit plots constructed from appropriate dose–response data. As discussed in the previous chapter (*6*), both LD_{50} and LD_{90} values for members of the TFMS herbicidal series were separately correlated with π and σ using the Hansch equation and its various modifications presented in this work. Except for obvious variations in the π and σ parameter coefficients necessary to account for the fact that the LD_{50} values for all derivatives are less than their corresponding LD_{90} values, the *same* over-all conclusions were predicted whether LD_{50} or LD_{90} values were used in carrying out the various structure–activity correlations. Because the LD_{90}value is a more meaningful designation of activity than the LD_{50} value in herbicidal studies, we report all correlation results and conclusions in terms of LD_{90}.

Surfactant. Polyoxyethylene (20) sorbitan monooleate (also known as Tween 80, At-Plus 109, or Polysorbate 80) was obtained from Atlas Chemical Industries and used as received. It was used at the 0.1% (w/v) concentration level in all herbicidal activity evaluations and partitioning studies requiring the use of surfactant.

Partition Coefficient. Partition coefficients of the TFMS herbicides in both the presence and absence of surfactant were determined between 1-octanol and pH 1.0 water (made acid by addition of $HClO_4$) by ultraviolet spectroscopy. The absorption spectrum of Tween 80 did not interfere with the spectra of the sulfonanilides (*6*).

Hammett Sigma Constant. Hammett sigma constants, which are a measure of the electron-donating and withdrawing capability of aromatic substituents, were taken from tabulations of Jaffe (*8*).

Results

Structure–activity correlations were carried out using least-squares regression analysis techniques on an IBM 360 computer. As in the accompanying publication (*6*), the data in Tables I and II were fitted to Equation 3 in stepwise fashion. Standard statistical tests were carried out at each stage of fitting to determine the over-all goodness of fit of the χ and σ data to the various equational forms examined. As in our previous study (*6*), the most statistically significant correlations were always obtained when activity data for meta-substituted and para-substituted TFMS herbicides were divided into two discrete series and fitted separately.

Correlation between σ and χ. Preliminary plots of the herbicidal partitioning data obtained in the presence and absence of Tween 80 indicated that the surfactant-induced change in partition coefficient (χ) and the Hammett sigma constant (σ—*see* Table I) were not independent variables. A particularly strong relationship existed, for example, between the σ_{para} values characterizing the 4-substituted TFMS compounds and χ. Empirically correlating σ with both first- and second-order terms in χ using stepwise regression methods yielded the following equations for the meta- and para-substituted series members.

Only para-substituted TFMS series members (n = 7):

$$\sigma_{para} = \underset{(\pm 0.196)}{0.062}\ \chi^2 + 0.087 \qquad (4a)$$

$$r = 0.140$$

$$r^2 = 0.020$$

$$SE = \pm 0.402$$

$$F = 0.10$$

$$\sigma_{para} = \underset{(\pm 0.273)}{1.661}\ \chi^2 - \underset{(\pm 0.500)}{3.024}\ \chi + 1.228 \qquad (4b)$$

$$r = 0.950$$

$$r^2 = 0.903$$

$$SE = \pm 0.141$$

$$F = 18.68$$

Only meta-substituted TFMS series members (n = 6):

$$\sigma_{meta} = \underset{(\pm 0.172)}{0.063}\ \chi^2 + 0.282 \qquad (5a)$$

$$r = 0.179$$

$$r^2 = 0.032$$

$$SE = \pm 0.246$$

$$F = 0.13$$

$$\sigma_{meta} = \underset{(\pm 0.775)}{1.175}\ \chi^2 - \underset{(\pm 1.197)}{1.753}\ \chi + 0.734 \qquad (5b)$$

$$r = 0.660$$

$$r^2 = 0.436$$

$$SE = \pm 0.217$$

$$F = 1.16$$

In Equations 4 and 5, r is the multiple correlation coefficient, r^2 is the "percent correlation," SE is the standard error of the equation (*i.e.*, the error in the calculated σ values), and F is the ratio of the mean sum of error squares removed by regression to the mean sum of squares of the error residuals not removed by regression. The F-values were routinely used in statistical tests to determine the goodness of fit of the above and following equations. The numbers in parentheses beneath the fit parameters in each equation denote the standard error in the respective pa-

rameters. They can also be used to judge the goodness of fit of a particular equation to the experimental data.

Referring to Equation 4b, it is evident that there is a statistically significant parabolic relationship between σ_{para} and χ for the 4-substituted TFMS derivatives. Both χ^2 and χ terms are necessary in the optimum correlation equation, as evidenced by the fact that the percent correlation (r^2) jumps from 2% in Equation 4a to 90% in Equation 4b upon stepwise addition of the linear term in χ. No simple linear correlation between σ_{para} and χ was found. Fits of the form $\sigma_{para} = a\chi + b$, for example, were even less statistically significant than the poorly correlated single-term quadratic relationship of Equation 4a.

As was true for σ_{para}, no statistically significant linear relationship between σ_{meta} and χ was found. Quadratic relationships for σ_{meta} similar in form to the relationships of Equations 4a and 4b are presented in Equations 5a and 5b for the meta-substituted TFMS derivatives. Both χ^2 and χ terms are again necessary for optimum correlation although the final correlation relationship in Equation 5b for σ_{meta} is not nearly as statistically significant ($r^2 = 0.44$) as the final relationship for σ_{para} in Equation 4b ($r^2 = 0.90$).

Although Equations 4b and 5b are difficult to interpret in a mechanistic sense, they do suggest that the change in partition coefficient ($\chi = \log P - \log P'$) to be expected upon addition of 0.1% Tween 80 to a herbicidal TFMS formulation will be influenced at least to some extent by three factors:

(1) The nature of the substituent (X) in the TFMS aromatic ring

(2) The position of substitution (meta or para) in the aromatic ring

(3) The electronegativity of the substituent (as reflected by the value of its Hammett sigma constant).

With respect to the above, it is not unlikely that the octanol/water partitioning behavior of all TFMS series members is influenced by the effects that the meta and para ring substituents exert on the acidic ($pK_a = 4.45$) parent $-NHSO_2CF_3$ side chain. Because of the direct resonance effects that they exert on the $-NHSO_2CF_3$ side chain, electron-withdrawing (positive σ) or electron-donating (negative σ) groups substituted in the para ring position of the TFMS parent compound, for example, might be expected to influence much more strongly the acidity (and hence the partitioning behavior) of the resulting TFMS derivative than would be the case for similarly substituted meta derivatives (where only inductive or field effects are possible). Furthermore, the significant correlation between σ_{para}, χ^2, and χ in Equation 4b strongly suggests that these resonance effects present in para-substituted derivatives probably influence the nature of the interaction (micelle formation, etc.) between herbicide and surfactant to a much greater degree than in the case of

meta-substituted TFMS derivatives (where the correlation between σ_{meta}, χ^2, and χ is poorer).

As shown below, Tween 80 addition to both meta- and para-substituted TFMS compounds often affects their herbicidal activity. This surfactant-induced change in activity (whether leading to ultimate enhancement or inhibition) in all cases depends significantly on the change in partition coefficient (χ) of the herbicide produced by surfactant addition. Obviously, the ability to predict χ from a simple knowledge of the Hammett σ constant characterizing a particular TFMS derivative (without resorting to time-consuming partitioning measurements) would in many cases prove advantageous. In this connection, the statistically significant Equation 4b can be rearranged and solved for χ using the quadratic equation to obtain the following expression.

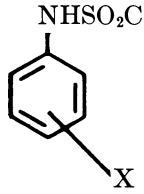
$$\chi_{para} = \frac{3.024 \pm \sqrt{9.145 - 6.644\,(1.228 - \sigma_{para})}}{3.322} = \frac{3.024 \pm \sqrt{0.986 + 6.644\,\sigma_{para}}}{3.322} \quad (6)$$

Equation 6 is very useful for estimating changes in the partition coefficient as well as changes in the herbicidal activity of para-substituted TFMS compounds (*see below*) produced when 0.1% Tween 80 is added to the herbicidal formulations. An expression analogous to Equation 6 can be derived from Equation 5b for meta-substituted TFMS series members although the predictive utility of this latter equation will obviously be limited by the lesser degree of correlation between σ_{meta}, χ^2, and χ.

Herbicidal Activity Correlations. Tables I and II give pre-emergence herbicidal activity and partition coefficient data gathered in the presence of 0.1% Tween 80 for the 15 TFMS compound evaluated in this study. For reasons discussed previously (*6*), in the correlations which follow, the Hammett sigma constant was assumed to be relatively unaffected by the presence of the surfactant, so that the σ-values listed in Tables I and II could be used to correlate data obtained both in the presence and absence of surfactant. Pertinent herbicidal activity data for the TFMS compounds acting on Foxtail grass are presented in Table I. Similar data for the same compounds acting on the broadleaf Wild Mustard are tabulated in Table II.

For reasons outlined in the previous chapter (*6*), TFMS compounds were separated into meta- and para-substituted series members, and structure–activity data for each grouping were correlated separately. Since our previous analysis (*6*) indicated that the 3-SCH_3-TFMS and 4-SCH_3-TFMS derivatives were probably oxidized *in vivo* to corresponding sulfoxide and/or sulfone derivatives, data points for these series

Table I. Pre-Emergence Control of Foxtail by Trifluoromethanesulfonanilides



Substituent, X	*LD_{90}, No Surfactant*	*LD_{90}', 0.1% Tween 80*	χ = [a] *Log P − Log P'*	σ	*log* $\left(\frac{LD_{90}'}{LD_{90}}\right)$
4 CF_3	9.70	10.6	1.57	+0.551	+0.039
4 Cl	6.40	8.15	1.36	+0.227	+0.105
4 SCH_3 [b]	5.10	5.10	1.04	−0.047	0.000
4 CH_3	82.0	40.0	1.10	−0.170	−0.312
4 F	8.0	10.1	0.75	+0.062	+0.101
4 OCH_3	58.5	23.0	0.70	−0.268	−0.405
4 SO_2CH_3	2.42	2.64	0.19	+0.728	+0.038
3 CF_3	15.0	21.5	1.30	+0.415	+0.156
3 Cl	7.8	14.4	—	+0.373	+0.266
3 F	8.35	18.2	0.62	+0.337	+0.338
3 SCH_3 [b]	—	2.06	0.35	+0.144	—
3 $COCH_3$	14.6	12.4	0.28	+0.306	−0.071
3 OH	128.0	204.0	0.41	−0.002	+0.202
3 SO_2CH_3	1.5	1.17	0.15	+0.647	−0.108
H [b]	8.0	12.7	0.45	0.000	+0.201

[a] *See* Ref. *6* for absolute values of log *P* and log *P'*.
[b] Omitted from correlations; *see* text and Ref. *6* for details.

members were again omitted from their respective groupings in deriving the final correlation equations. Computerized regression analysis techniques were used in the usual manner to fit the TFMS structure–activity data in the separate meta and para groupings to Equation 3. During the fitting procedure, χ^2, χ, and σ terms were added in stepwise fashion, with the order of inclusion of these terms determined on the basis of statistical *F*-tests. The stepwise equations resulting from the regression analysis are listed below for both Foxtail and Wild Mustard.

Correlation of Foxtail Activity Data

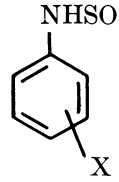
4-Substituted TFMS derivatives (n = *6*):

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.197)}{0.415\sigma - 0.151} \qquad (7a)$$

$$r = 0.726 \qquad SE = \pm 0.173$$

$$r^2 = 0.527 \qquad F = 4.46$$

Table II. Pre-Emergence Control of Wild Mustard by Trifluoromethanesulfonanilides



Substituent, X	*LD_{90}, No Surfactant*	*LD_{90}', 0.1% Tween 80*	*χ = [a] Log P − Log P'*	σ	*log $\left(\frac{LD_{90}'}{LD_{90}}\right)$*
4 CF_3	3.35	2.64	1.57	+0.551	−0.103
4 Cl	1.36	1.35	1.36	+0.227	−0.003
4 SCH_3 [b]	42.0	25.2	1.04	−0.047	−0.222
4 CH_3	18.9	10.5	1.10	−0.170	−0.255
4 F	1.85	2.31	0.75	+0.062	+0.096
4 OCH_3	35.6	80.5	0.70	−0.268	+0.354
4 SO_2CH_3	31.6	79.0	0.19	+0.728	+0.398
3 CF_3	6.15	1.95	1.30	+0.415	−0.499
3 Cl	5.20	5.20	—	+0.373	0.000
3 F	5.55	4.67	0.62	+0.337	−0.075
3 $COCH_3$	24.3	33.0	0.28	+0.306	+0.133
3 OH	46.2	32.9	0.41	−0.002	−0.147
3 SO_2CH_3	17.4	26.8	0.15	+0.647	+0.188
3 SCH_3 [b]	9.5	17.2	0.35	+0.144	+0.258
H [b]	4.6	6.4	0.45	0.000	+0.143

[a] *See* Ref. *6* for absolute values of log *P* and log *P'*.
[b] Omitted from correlations; *see* text and Ref. *6* for details.

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.173)}{0.085\,\chi} + \underset{(\pm 0.219)}{0.424\,\sigma} - 0.232 \qquad (7b)$$

$$r = 0.750$$

$$r^2 = 0.562$$

$$SE = \pm 0.193$$

$$F = 1.92$$

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.506)}{-2.122\,\chi^2} + \underset{(\pm 0.922)}{3.944\,\chi} + \underset{(\pm 0.280)}{1.542\,\sigma} - 1.751 \qquad (7c)$$

$$r = 0.977$$

$$r^2 = 0.955$$

$$SE = \pm 0.075$$

$$F = 14.26$$

$$\chi_{max} = +0.929$$

3-Substituted TFMS derivatives (n = 5):

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.207)}{0.212\,\chi} - 0.014 \tag{8a}$$

$$r = 0.508$$

$$r^2 = 0.258$$

$$SE = \pm 0.188$$

$$F = 1.04$$

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.289)}{-1.098\,\chi^2} + \underset{(\pm 0.444)}{1.868\,\chi} - 0.413 \tag{8b}$$

$$r = 0.954$$

$$r^2 = 0.910$$

$$SE = \pm 0.080$$

$$F = 10.11$$

$$\chi_{max} = +0.851$$

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.536)}{-1.158\,\chi^2} + \underset{(\pm 0.820)}{1.958\,\chi} + \underset{(\pm 0.319)}{0.054\,\sigma} - 0.453 \tag{8c}$$

$$r = 0.955$$

$$r^2 = 0.912$$

$$SE = \pm 0.112$$

$$F = 3.47$$

$$\chi_{max} = +0.845$$

Comparison of Equations 7a to 7c for the 4-substituted TFMS derivatives acting on Foxtail grass clearly demonstrates that χ^2, χ, and σ terms are all statistically significant in the final correlation relationship of Equation 7c. The "change-in-activity" surface represented by Equation 7c exhibits a maximum at $\chi_{max} = +0.93$. This maximum corresponds to a ridge along the parabolic log (LD_{90}'/LD_{90}) envelope and represents the locus of 4-substituted TFMS compounds for which the addition of 0.1% Tween 80 causes the greatest inhibitory effects on herbicidal activity arising only from surfactant-induced changes in the partition coefficient (*i.e.*, χ). Substituting $\chi_{max} = +0.93$ into Equation 4b one is able to estimate that a TFMS derivative containing a para-substituted side chain characterized by a Hammett sigma constant of ≈ -0.15 would lie on or near the inhibitory χ_{max} ridge. Referring back to Equation 7c,

however, we note that although surfactant-induced changes in log P should be highly unfavorable from an activity standpoint for a TFMS derivative containing a para substituent with $\sigma_{para} = -0.15$, the fact that the coefficient of σ in Equation 7c is positive also implies that the over-all activity of the herbicide might still remain acceptable because of the favorable negative contribution of the σ term to log (LD_{90}'/LD_{90}).

Similar comparison of Equations 8a to 8c for 3-substituted TFMS derivatives acting on Foxtail grass reveals that only χ^2 and χ terms are statistically significant as shown in Equation 8b. This implies that for meta-substituted derivatives surfactant-induced changes in the partition coefficient are primarily responsible for changes in the herbicidal activity of the various series members. The fact that the Hammett sigma constant is poorly correlated with surfactant-induced changes in herbicidal activity for meta-substituted TFMS derivatives but strongly correlated with surfactant-induced activity variations for para-substituted derivatives would appear to be predictable from the relationships of Equations 4b and 5b. In Equation 5b, for example, it was shown that surfactant-induced changes in log P for meta series members are not strongly dependent on the value of the Hammett sigma constant. Since herbicidal activity variations produced by the addition of Tween 80 depend primarily on corresponding partition coefficient changes in the 3-substituted TFMS series members, it follows from Equation 5b that one would have expected little or no contribution of the σ term in Equation 3 to the final regression relationship for meta-TFMS compounds. The essential absence of surfactant effects on σ for 3-substituted TFMS series members acting on Foxtail can be readily visualized by comparing Figures 2*b* and 2*d* in the previous chapter (*6*).

Correlation of Wild Mustard Activity Data

Proceeding in a manner identical to that outlined above for Foxtail grass, the following stepwise regression relationships were derived for the broadleaf Wild Mustard.

4-Substituted TFMS derivatives (n = *6*):

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.155)}{-0.407\,\chi + 0.466} \qquad (9a)$$

$$r = 0.795$$

$$r^2 = 0.633$$

$$SE = \pm 0.174$$

$$F = 6.89$$

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.396)}{0.207}\,\chi^2 - \underset{(\pm 0.726)}{0.776}\,\chi + 0.587 \qquad (9b)$$

$$r = 0.814$$

$$r^2 = 0.663$$

$$SE = \pm 0.192$$

$$F = 2.95$$

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 1.526)}{0.754}\,\chi^2 - \underset{(\pm 2.784)}{1.773}\,\chi - \underset{(\pm 0.845)}{0.318}\,\sigma + 0.986 \qquad (9c)$$

$$r = 0.828$$

$$r^2 = 0.685$$

$$SE = \pm 0.228$$

$$F = 1.45$$

3-Substituted TFMS derivatives (n = *5*):

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.108)}{-0.572}\,\chi + 0.236 \qquad (10a)$$

$$r = 0.951$$

$$r^2 = 0.903$$

$$SE = \pm 0.098$$

$$F = 28.05$$

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.084)}{-0.569}\,\chi + \underset{(\pm 0.164)}{0.280}\,\sigma + 0.137 \qquad (10b)$$

$$r = 0.980$$

$$r^2 = 0.961$$

$$SE = \pm 0.076$$

$$F = 24.48$$

$$\log\left(\frac{LD_{90}'}{LD_{90}}\right) = \underset{(\pm 0.449)}{-0.259}\,\chi^2 - \underset{(\pm 0.686)}{0.177}\,\chi + \underset{(\pm 0.267)}{0.382}\,\sigma + 0.009 \qquad (10c)$$

$$r = 0.985$$

$$r^2 = 0.971$$

$$SE = \pm 0.094$$

$$F = 10.99$$

When Equations 9a to 9c for the 4-substituted TFMS derivatives acting on Wild Mustard are compared, it is evident on the basis of the F-values for each equation that only a single term in χ is statistically significant and that Equation 9a provides the best mathematical description of surfactant effects on herbicidal activity. Essentially the same conclusion holds true for the regression relationships in Equations 10a to 10c for the 3-substituted TFMS compounds. Again, a single term in χ (*see* Equation 10a) provides a 90% correlation of the structure–activity data. At least to a first approximation, therefore, it would appear that for both meta- and para-substituted TFMS compounds acting on the broadleaf Wild Mustard, the observed changes in herbicidal activity caused by Tween 80 addition to herbicidal formulations can be explained in terms of the corresponding changes the surfactant induces in the partition coefficients of the various derivatives. Equations 9a and 10a both suggest that those TFMS derivatives experiencing the greatest decrease in octanol/water partition coefficient in the presence of 0.1% Tween 80 ($\chi = [\log P - \log P']$ large and positive) will also exhibit the greatest enhancement in herbicidal activity when surfactant is added.

Discussion

Surface-active agents can alter the effectiveness of herbicides in various ways (*6*, *9*). By limiting activity evaluations to pre-emergence studies, one can eliminate such difficult-to-control variables as bouncing, coverage, evaporation, and wetting. The herbicidal system examined here consisted of a series of meta- and para-substituted trifluoromethanesulfonanilide (TFMS) pre-emergence herbicides whose activity was evaluated in the presence and absence of a constant concentration (0.1% w/v) of the nonionic surfactant Tween 80. Admixture of this penetrant aid with the various members of the TFMS herbicidal series led to enhancement of activity in some cases and inhibition of activity in others (*6*). Still other series members were relatively unaffected by surfactant in their activity towards Foxtail grass and the broadleaf Wild Mustard (*cf.* LD_{90} and LD_{90}' values in Tables I and II).

As judged from octanol/water partitioning experiments in the presence and absence of Tween 80, the principal effect of adding surfactant to the TFMS herbicides at the 0.1% (w/v) level was to alter their relative solubilities in the aqueous and lipid phases. Since the trifluoromethanesulfonanilides are all quite hydrophobic, the addition of Tween 80 in every case increased the water solubility of the compound in question and hence decreased its octanol/water partition coefficient (*6*).

When the most statistically significant relationships derived for the meta- and para-TFMS herbicides acting on Foxtail and Wild Mustard

are compared (*i.e.*, Equations 7c, 8b, 9a, and 10a), it is evident that herbicidal activity variations (whether enhancement or inhibition) caused by Tween 80 addition can be traced almost entirely to corresponding surfactant-induced changes in the lipid–water partitioning behavior of the active compounds (*i.e.*, variations in $\chi = \log P - \log P'$). This is certainly true for meta and para TFMS derivatives acting on Wild Mustard (*see* Equations 9a and 10a) and for the meta TFMS herbicides acting on Foxtail (Equation 8b). In fact, only for the para-substituted TFMS derivatives acting on Foxtail (Equation 7c) does the Hammett σ constant make a significant additional contribution to the over-all correlation relationship. The fact that σ_{para} and χ are also strongly correlated (*see* Equation 4b) further illustrates the extreme importance of surfactant-induced effects on the partition coefficient of each TFMS compound in determining its ultimate herbicidal activity. This strong correlation between σ_{para} (and to a lesser extent σ_{meta}) and χ unfortunately implies, however, that these variables cannot be altered totally independently of each other if attempts are made to change TFMS herbicidal activity by surfactant addition and/or substituent variations in the parent molecule. On the other hand, the large positive value of the σ parameter coefficient in Equation 7c suggests that TFMS compounds substituted with strongly electron-donating substituents (negative σ values) in the para aromatic ring position will exhibit enhanced activity against Foxtail when Tween 80 is added to the herbicidal formulations at a constant concentration level (*cf.* in Table I, for example, 4 CH_3–TFMS ($\sigma_{para} = -0.170$) and 4 OCH_3–TFMS ($\sigma_{para} = -0.268$) activity against Foxtail in the presence and absence of surfactant.

Although experiments in this study were carried out at a single concentration level of a specific surfactant and as such are limited in scope, it is interesting to speculate on the implications of our results, particularly if we extrapolate them to situations where the concentration and type of surfactant are varied. These speculations must, of course, be verified by further experimentation.

In this regard, Jansen (*1*) has suggested that through proper use of surfactants herbicidal spray solutions can be tailored to meet specific situations. Steffens and Cathey (*10*) have likewise pointed out that for enhancement of herbicidal activity both the relative concentrations of the herbicide and an added surfactant are critical. Our results tend to substantiate these earlier conclusions. The fact that log (LD_{90}'/LD_{90}) in Equation 3 exhibits χ-dependencies which are different for Foxtail and Wild Mustard suggests the possibility of altering grass and broadleaf selectivity of the TFMS herbicides by proper use and application of surfactants. Considering first the meta-substituted TFMS compounds, we

note that if one wished to increase broadleaf (Wild Mustard) selectivity over grasses (Foxtail), Equations 8b and 10a suggest that this might be achieved by adding a particular type or amount of surfactant until ($\log P - \log P'$) equals +0.85 where meta-TFMS activity *vs.* Foxtail is minimized. If one wished to increase grass control while decreasing activity against broadleafs, Equations 8b and 10a suggest that this might be achieved by adding a surfactant that increases the partition coefficient of the meta-TFMS herbicides ($\log P - \log P'$ positive). Extending this line of reasoning to the para-TFMS herbicides, we note that an increase in grass control over broadleaf weed control might be obtained by increasing log P through appropriate surfactant addition such that ($\log P - \log P'$) is approximately equal to −2.0 (*cf.* signs of χ dependencies in Equations 7c and 9a). Increasingly selective control of broadleaf weeds over grasses might in turn be achieved by minimizing grass control through addition of a surfactant that would adjust ($\log P - \log P'$) to +0.93. At this value of χ, grass control is minimized, and broadleaf control is about the same as if no surfactant were added (*cf.* Equations 7c and 9a).

As pointed out in the previous chapter (*6*), our analysis has demonstrated that surfactant addition to herbicidal formulations can produce a variety of activity-enhancing and inhibitory effects. These effects appear to be caused primarily by alterations in the lipid–water partitioning behavior of the herbicides caused by surfactant addition. Indiscriminate use of surfactants as penetrant aids and formulating agents in herbicidal screening and field tests should thus be avoided. Herbicidal evaluations in the presence and absence of surfactant coupled with regression analyses of the type discussed in this and the previous chapter (*6*) should provide valuable clues regarding the most advantageous use of surfactants in enhancing over-all herbicidal activity and selectivity.

Acknowledgments

The authors are indebted to J. Belisle for the partition coefficient measurements and to J. Waddington and D. Pauly for the pre-emergence herbicidal activity evaluations.

Literature Cited

1. Jansen, L. L., Gentner, W. A., Shaw, W. C., *Weeds* (1961) **9**, 381.
2. Becher, P., Becher, D., ADVAN. CHEM. SER. (1969) **86**, 15.
3. Becher, P., "Emulsions: Theory and Practice," 2nd ed., pp. 232, Reinhold, New York, 1965.
4. Foy, C. L., Smith, L. W., ADVAN. CHEM. SER. (1969) **86**, 55.
5. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **85**, 1616.

6. Yapel, A. F., ADVAN. CHEM. SER. (1972) **114**, 183.
7. Trepka, R., Harrington, J. K., Robertson, J. E., Waddington, J. T., *J. Agr. Food Chem.* (1970) **18**, 1176.
8. Jaffe, H. H., *Chem. Rev.* (1953) **53**, 191.
9. Holly, K., "The Physiology and Biochemistry of Herbicides," L. J. Audus, Ed., Academic, New York, 1964.
10. Steffens, G. L., Cathey, H. M., *J. Agr. Food Chem.* (1969) **17**, 312.

RECEIVED February 10, 1972.

14

The Hansch Structure–Activity Approach As an Aid in Designing New Biologically Active Chemicals

W. BROCK NEELY

Chemical Biology Research, The Dow Chemical Co., Midland, Mich. 48640

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

To examine the advantages and disadvantages of the Hansch multiple regression analysis in structure–activity studies, a series of compounds were prepared for their insecticidal, herbicidal, and analgesic activity. For the insecticidal phosphoramidate group there seems to be no relationship between cholinsterase inhibition and insecticidal activity. One group of compounds showed good correlation between insecticidal activity and partition coefficients and hydrolysis constants. In the herbicidal series of compounds, R_1 in the common structure seems to play an important role in producing a drug with the proper hydrophile/lipophile balance. Imadozolines synthesized for their analgesic characteristics were correlated using both partition coefficient and the HOMO index. Inclusion of the latter term is significant at the 5% level.

This article describes our application of the Hansch technique to various structure–activity problems. Thus, it is not an exhaustive review of the literature but a discussion of the problems and pitfalls encountered in the use of such correlation studies. For a more thorough review the reader is referred to other articles (*1–3*).

Historical Aspects

During the 19th century many natural products were isolated and purified, and their chemical structures were determined. Some of them exhibited pharmacological activity, and it soon became apparent that the activity could be either enhanced or lowered by changing the functional groups attached to the parent structure. Thus, modifications of structure

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

became common and are still important. It is hoped that such changes will widen the differences between toxic and therapeutic doses. Furthermore, by alteration in structure new properties may come to light—properties that had been hidden because of previous high toxicity. Some people have referred to this approach in a facetious manner as "methyl, ethyl, butyl, and futile." Far from being futile, it has resulted in scores of useful compounds which possess high intrinsic activity with minimum side effects.

Ideas on structure–activity relations became more quantitative with the publication of the work of Hansch (*4*) and Free and Wilson (*5*) in 1964. Since that time, many investigators have examined structures using these more rational approaches to design active chemicals. It is still hoped that a unified theory will evolve in which structures for a specific biological activity may be predicted. However, concepts generated in one series of drugs cannot be used indiscriminately in a totally different series. Hansch, in working out his model for correlation studies, has had a great impact in this field. With the many correlations he has made, he is now in a position to begin looking for such unified theories. A recent publication (*6*) indicates the type of analysis that he is now performing.

Our use of the Hansch technique has been pragmatic—*i.e.*, to uncover the best drug in the shortest and in the most efficient way. We felt that by applying the concepts generated by Hansch we could discover relationships between structures and activity which would allow us to complete the synthesis faster. During these investigations, certain facts became obvious.

(1) In examining an extended series of chemicals, the busy chemist had made most of the possible structures. Thus, the resulting analysis, while intellectually satisfying, was always after the fact.

(2) In cases where predictions could be made, the predicted structures were usually too difficult to make and hence too costly.

(3) The most useful approach was to consider a limited number of structures where an intensive synthesis program was not underway and to point out trends. This limitation usually resulted in an analysis that was close to being statistically significant.

(4) The analysis depends on good biological data, and this is sometimes hard to come by. The information that was usually missing was data on the inactive structures. It is impossible to handle, information in a quantitative manner, that reads "greater than 100 ppm."

(5) After assuming a certain mechanism for an observed biological activity—*e.g.* inhibition of cholinesterase and insecticidal activity—it was interesting to look for deviations. Such deviations could point to new mechanisms and hence a new synthesis scheme.

(6) If no mechanism existed, the analysis could help predict what a possible mechanism might be.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

In Biological Correlations—The Hansch Approach; Van Valkenburg, W.;
Advances in Chemistry; American Chemical Society: Washington, DC, 1974.

(7) Finally, and perhaps most importantly, this type of program for maximum utility should be either carried on by the synthesis chemist himself or at least synthesis capabilities should be readily available.

Our Approach

Our approach is similar to the system devised by Hansch. The model assumes that the amount of externally applied drug found at the receptor site is related to the partition coefficient. Once the drug reaches the receptor site, the intrinsic activity of the molecules takes over and triggers a reaction which elicits the biological response. The model is illustrated by Equation 1.

$$\log (BR) = a\pi^2 + b\pi + c \log K_x + d \qquad (1)$$

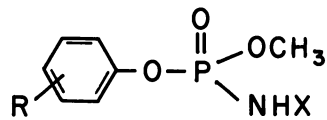
where

BR = biological response
π = log (Px/P_H)
P_x = partition coefficient of substituted compound
P_H = partition coefficient of parent compound
K_x = reactivity index
a,b,c,d = equation constants

There are many ways to measure the reactivity index, and all of them are feasible in such a study. By finding one that works—*i.e.*, gives significant statistics in Equation 1—statements may be formulated regarding possible mechanisms. In this type of work, it must be remembered that it is not valid to imply causality to a correlation. However, from a pragmatic point of view, it is possible to set up models that can be tested by further synthesis. The measurements used for the reactivity index in the following examples include hydrolysis constants and molecular orbital calculations.

Insecticides

A series of organophosphates represented by Structure I were examined. Such chemicals have been used for many years to control insects.



I

Classically, the mechanism used to depict this biological activity is shown in Figure 1 where the primary attack is the phosphorylation of the

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

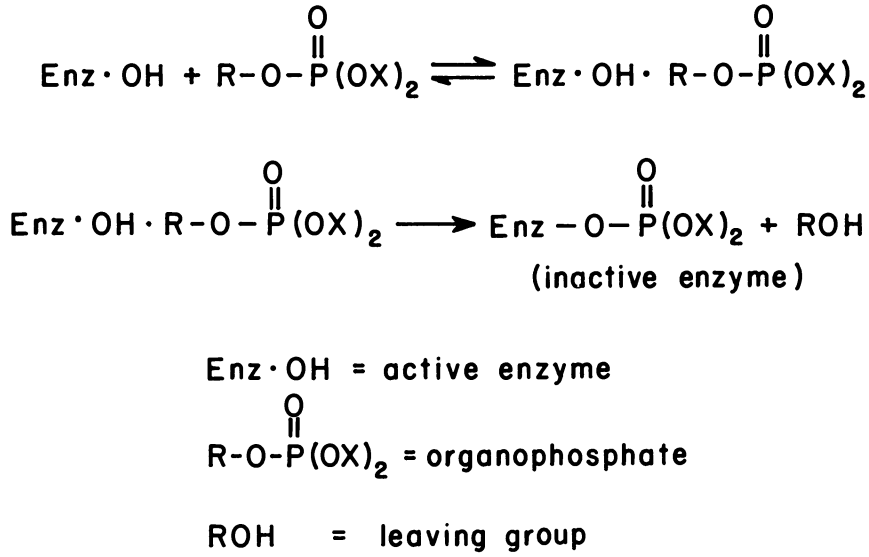


Figure 1. Model for the inhibition of cholinesterase by the attack of an organophosphate

esteratic site in cholinesterase. From this mechanism it is deduced that those phosphates which are most susceptible to nucleophilic attack will be the most potent insecticides. If this mechanism is valid, the rate of hydrolysis of the phosphate by OH^- may be used as a measure of the reactivity. The phosphoramidates used along with the biological data, partition coefficients, and hydrolysis rates are shown in Table I (7). Using multiple regression techniques, these data were fitted to Equation 1 and generated Equation 2 (*see* Figure 2):

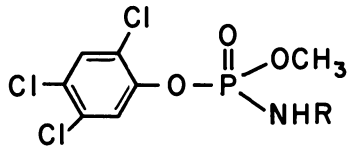
$$\log (LD_{50}) = 0.941\pi - 6.03\pi^2 - 0.509 \log k + 8.44 \qquad (2)$$

where the square of the correlation coefficient was 0.88 and the standard deviation was 0.116.

Equation 2 should be suitable for predicting LD_{50} values in a series of phosphoramidates. The only criterion that must be met is that all the parameters must have a common base. The substituted phenyl *O*-methyl methyl phosphoramidates in Table II meet this requirement. Consequently, it is possible to use Equation 2 to calculate an LD_{50} value for each member of the series. The results indicated a complete lack of agreement between the experimental and calculated LD_{50}—*e.g.*, the 4-Cl derivative in Table II has an experimental value of 0.26; the calculated value was 330. This is surprising and might mean that the proposed mechanism of action does not hold for this particular series. This is again

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

Table I. Chemical and Biological Data for the 2,4,5-Trichlorophenyl O-Methyl Phosphoramidates (II)



II

R	*Topical*[a] μg/*fly* LD_{50}	π[b]	k (min^{-1})[c]
H	0.084	2.04	60.5
Methyl	0.076	2.56	3.45
Ethyl	0.05	3.07	2.24
n-Propyl	0.067	3.60	2.10
Isopropyl	0.036	3.42	1.47
n-Butyl	0.166	4.04	1.96
tert-Butyl	0.35	3.72	0.10
sec-Butyl	0.152	3.86	1.43
Isobutyl	0.18	3.99	1.72

[a] Represents the LD_{50} of the toxicant in mg/fly.
[b] Partition coefficient determined by procedure of Hansch (*11*).
[c] Rate of hydrolysis in 0.1*N* NaOH.

reflected in the fact that no correlation is observed in trying to fit the LD_{50} values of Table II with Equation 1.

By contrast, the enzyme-inhibition data of Table II correlated quite well with the partition and hydrolysis data (Equation 3).

$$\log K = 0.905\pi + 1.17 \log K + 3.21 \quad (3)$$

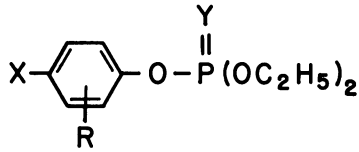
where $r^2 = 0.92$ and standard deviation $= 0.315$

This would again suggest the possibility that for these phosphoramidates there is no relation between cholinesterase inhibition and insecticidal activity. We are investigating other structural features in this series to see if some alternative insecticidal mechanism may be acting.

Another series of phosphates that was examined is shown in Structures IV to VI. That analysis (*8*) was done to determine the nature of the contribution that the thiomethyl and the sulfonylmethyl groups had on insecticidal activity. Briefly, the results are:

(1) With the group represented by Structure V, good correlation was found between insecticidal activity (LD_{50} on flies) and the partition coefficients and hydrolysis constants (Equation 4).

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014



IV, X = SCH_3 ; Y=O

V, X = CH_3SO_2; Y=O

VI, X= SCH_3 ; Y=S

$$\log LD_{50} = 0.84\pi^2 - 0.194\pi - 0.626 \log k - 1.76 \quad (4)$$

The resulting high correlation ($r^2 = 0.98$) may be the result of the fact that this group of materials represents a situation where all pertinent transformations have occurred—*i.e.*, conversion of SCH_3 to CH_3SO_2 and P═S to P═O. Consequently, the group V structures are a series where

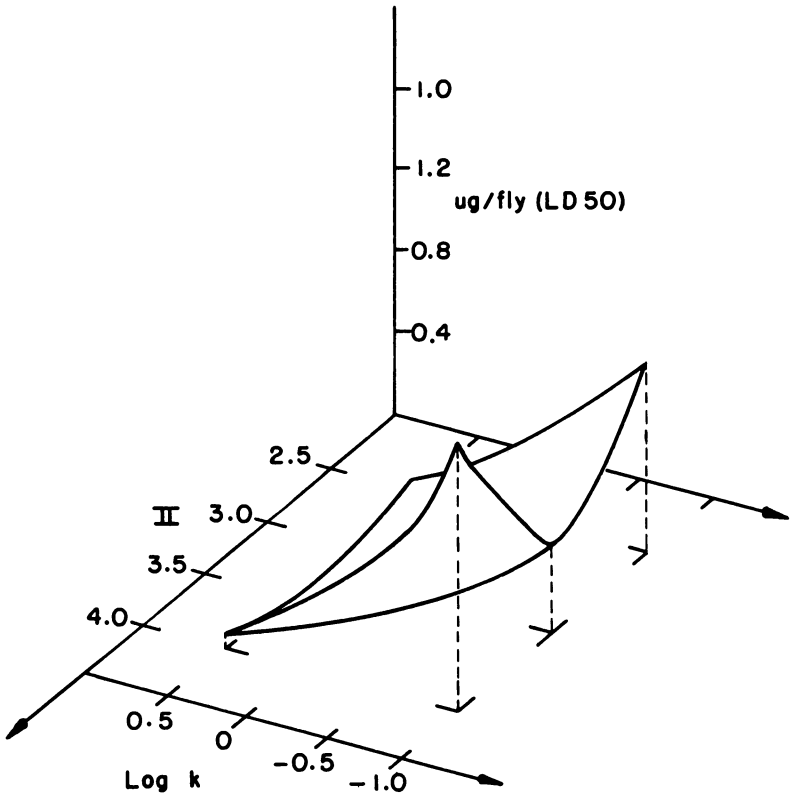
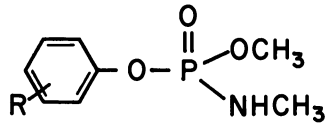


Figure 2. Graphical representation of Equation 2

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

Table II. Chemical and Biological Data for Substituted Phenyl O-Methyl Methyl Phosphoramidates (III)



III

R	*Topical μg/fly LD_{50}*	π [a]	K (min^{-1}) [b]	K [c]
4-OCH_3	>10	0.56	0.12	4.28×10^{-2}
4-*tert*-Butyl	>10	2.22	0.078	7.83×10^{-3}
3-*tert*-Butyl	>10	2.14	0.12	7.11×10^{-3}
4-Cl	0.26	1.28	0.36	6.02×10^{-3}
2, 4-diCl	0.1	1.86	1.87	1.07×10^{-5}
2-Cl	3.0	1.11	1.00	7.40×10^{-3}
2Cl, 4-*tert*-Butyl	0.48	2.87	0.608	1.10×10^{-6}
H	>10	0.52	0.12	6.94×10^{-2}
2, 4, 5-triCl	0.076	2.56	3.45	1.07×10^{-6}

[a] Determined by the Hansch procedure (*11*).
[b] Rate of hydrolysis in 0.1*N* NaOH.
[c] Bimolecular rate constant for inhibition of fly head cholinesterase.

the measured parameters are relevant as far as insecticidal activity is concerned.

(2) The group represented by Structure IV could not be fitted to the Hansch equation with the existing parameters. This might be explained if the conversion of the thiomethyl to sulfonylmethyl were rate limiting. Perhaps a significant correlation would be found if the rate of this conversion were used in the regression rather than the hydrolysis constant.

(3) Compounds in series VI were interesting, in that the correlation that was found (Equation 5) predicts that the compounds which are most stable to hydrolysis will have the highest activity.

$$\log LD_{50} = 0.524 \log k + 0.045 \quad (5)$$

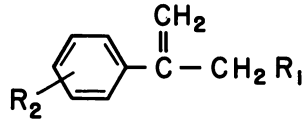
This series is an example where the analysis is close to being statistically significant. However, the prediction, since it is contrary to expectation, needs to be investigated.

Herbicides

Structure VII (with R = CCl_3 and R_2 = H) is useful for controlling water grass in rice paddies (*9*). We have done some studies on this and

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

related derivatives, as shown in Table III; Holmsen and Bidlack in our laboratory collected data on this series (*10*). The Hansch procedure on



VII

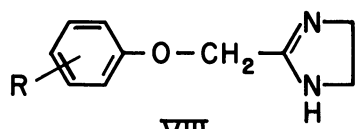
the data in Table III yielded Equation 6, which had a correlation coefficient of $r = 0.89$.

$$\log (\text{activity}) = 0.139\pi^2 + 0.50\pi + 1.538 \quad (6)$$

This is an interesting series in that R_1 is insulated from the ring; consequently, any electronic influence of the substituent will have negligible influence on the ring. Our tentative conclusion from the above analysis is that R_1 plays an important role in obtaining a drug that has the proper balance of hydrophobic and hydrophilic properties. The electronic effect of ring substituents was partly described in an earlier report (*12*).

Analgesics

This particular series of drugs represents a situation where the analysis suggests a possible mechanism of action. The chemicals to be discussed are a group of imidazolines, VIII, which were synthesized for their analgesic characteristics (*13*).



VIII

A reactivity index suitable for use in Equation 1 was calculated by using the simple molecular orbital techniques described by the Pullmans (*14*). Many indexes may be deduced from this type of procedure. The one that seemed to have the most significance for the correlation was the energy of the highest occupied molecular orbital (HOMO). This index is a relative measure of the ability of an electron to be transferred to an acceptor molecule. The calculations were performed on the substituted phenol in the imidazoline structure. This simplification was made since it could be assumed that any perturbation caused by the imidazole would be insulated from the rest of the molecule by the methylene group.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

Table III. Herbicidal and Partition Data for Substituted Styrene Derivatives (VII)

R_1	R_2	π [a]	*Activity* [b]
CH_3	—	0	
$CHCl_2$	—	1.20	100
CCl_3	—	1.80	95
CCl_3	4-CH_3	2.53	77
CCl_3	4-iso-propyl	3.10	73
CHCl C-OCH_3	—	0.48	52
CCl_2	—	3.00	49
CH_2OH	—	0.50	18

[a] π was based on calculated values as described by Hansch (*11*).
[b] Activity expressed as % control of grasses where soil was pretreated with chemical at the rate of 50 lb/acre.

Publication Date: August 1, 1974 | doi: 10.1021/ba-1972-0114.ch014

The data for this series of drugs are shown in Table IV. Multiple regression analysis generated the following equations:

$$\text{Log } ED_{50} = 0.445\pi + 1.015 \quad (7)$$
$$r^2 = 0.406$$

$$\text{Log } ED_{50} = 0.655\pi^2 - 1.00\pi + 0.455 \quad (8)$$
$$r^2 = 0.531$$

$$\text{Log } ED_{50} = 0.945\pi^2 - 1.850\pi + 7.90\,(\text{HOMO}) - 5.117 \quad (9)$$
$$r^2 = 0.944$$

The *F* test indicated that the inclusion of HOMO term in Equation 9 is significant at the 5% level.

The model that emerges from this study consists of a receptor site which has the ability to accept electrons from an electron-donating drug. The association is similar to a Michaelis-Menten type; hence, the biological reaction would have a relatively short half-life. In addition to these electronic considerations, a proper balance of hydrophobic and hydrophilic properties must be maintained.

Table IV. Biological and Chemical Parameters for Imidazoline Derivatives (VIII)

R	ED_{50} [a]	π [b]	*HOMO* [c]
2, 6-diCH_3	0.6	1.36	0.716
2-Br, 6-Cl	2.0	1.34	0.802
2, 6-diBr	3.2	1.50	0.797
2, 6-diCl	3.3	1.18	0.808
2, 6-diCl, 3-CH_3	5.4	1.69	0.789
2, 6-diOCH_3	25	−0.66	0.618

[a] Oral dose in mg/kg (HCl mouse writhing test).
[b] Determined from Hansch (*11*).
[c] The smaller the value, the greater the ease of electron donation.

Summary

Work is going on in molecular orbital calculations and in the role of solvent interactions (*15*). As theories become more refined, they will undoubtedly be incorporated into equations relating biological activity to structure. Such a situation will allow us to make more accurate predictions of what structures are needed for a specified type of activity. In addition, they will be of great assistance in learning more about the nature of the receptor site. A good example of this latter type of study is the work by Wohl (*16*) on the benzothiadiazene antihypertensive agents.

Literature Cited

1. Purcell, W. P., Singer, J. A., Sundaram, K., Parks, G. L., "Medicinal Chemistry," A. Burger, Ed., p. 164, Wiley, New York, 1970.
2. Hansch, C., *Accounts Chem. Res.* (1969) **2,** 232.
3. Cammarate, A., *J. Med. Chem.* (1969) **12,** 314.
4. Hansch, C., Fujita, T., *J. Amer. Chem. Soc.* (1964) **86,** 1616.
5. Free, S. M., Jr., Wilson, J. W., *J. Med. Chem.* (1964) **7,** 395.
6. Hansch, C., Lien, E. J., *J. Med. Chem.* (1971) **14,** 653.
7. Neely, W. B., Whitney, W. K., *J. Agr. Food Chem.* (1968) **16,** 571.
8. Neely, W. B., Allison, W. E., Crummett, W. B., Kauer, K., Reifschneider, W., *J. Agr. Food Chem.* (1970) **18,** 45.
9. DeDatta, S. K., Bantalan, R. T., Park, J. K., *Nature* (1960) **221,** 64.
10. Holmsen, T. W., Bidlack, H. D., unpublished data.
11. Fujita, T., Iwasa, J., Hansch, C., *J. Amer. Chem. Soc.* (1964) **86,** 5175.
12. Neely, W. B., in "Molecular Orbital Studies in Chemical Pharmacology," L. Kier, Ed., p. 121, Springer, New York, 1970.
13. Neely, W. B., White, H. C., Rudzik, A., *J. Pharm. Sci.* (1968) **57,** 1176.
14. Pullman, B., Pullman, A., "Quantum Biochemistry," Wiley, New York, 1963.
15. Cammarata, A., "Molecular Orbital Studies in Chemical Pharmacology," L. Kier, Ed., p. 156, Springer, New York, 1970.
16. Wohl, A. J., *Mol. Pharmacol.* (1970) **6,** 195.

RECEIVED May 10, 1971.

15

Molecular Orbital Studies of Biological Molecule Conformations

LEMONT B. KIER

Massachusetts College of Pharmacy, 179 Longwood Ave., Boston, Mass. 02115

Over the past five years a number of studies have been reported on the prediction of biological molecule conformation using molecular orbital theory. Several all-valence, semiempirical molecular orbital methods have been used in these studies, with the extended Hückel theory (EHT) being the most widely used to date. The agreement between EHT-predicted conformations of biological molecules and experimental evidence has been quite significant. A number of hypotheses of drug mechanisms have been proposed, based on these calculations. The utility of this approach and the potential it affords for new insight into biological structure and mechanism warrant its incorporation into the drug research armamentarium.

The conformation of a biologically active molecule very likely plays an important part in the total structural characteristics contributing to activity. Pronounced differences in biological potency in a closely related chemical series may well be the result of minor changes in the preferred conformation of these molecules. The result of these changes may be an incomplete or an inappropriate interaction with a biological receptor or enzyme. Thus, some optimum positioning of essential features in an active molecule must be necessary for biological efficacy. A knowledge of conformational preference is thus of vital concern.

Evaluation of Conformation

The prediction of the conformation of molecules has been of considerable interest to biological scientists for many years. Early approaches

centered on an intuition based on the presumed repulsive interaction of bulky groups across space. This gave rise to general rules of conformational preference which had some utility with hydrocarbons. Unfortunately, this intuition was unable to perceive the attractive forces which are also a part of the total influence on conformation. Attractive forces are particularly prominent in heteroatom molecules, and these are predominantly what the biological scientist encounters.

The advent of x-ray crystallography has permitted the mapping of the atoms of molecules in the solid state. The relevance of these conformations to solution phenomena is, however, obscure. In the crystal, the molecules are closely packed, interacting with each other and with gegenions if present. This is probably not the situation normally encountered in the dilute solutions of the biological milieux. Thus, biological conclusions derived from x-ray-derived conformations must always be considered in this light.

A more useful experimental approach to predicting conformation in a biological environment is through the use of NMR analysis in water. These data, if properly analyzed, give a time average conformation which can be of considerable value in subsequent biological interpretations. It is necessary, however, actually to have the compound under study, and frequently the analysis of the NMR data is extremely complex.

Other solution techniques for predicting preferred conformation include ORD, dipole moment and spectroscopic methods. Each is capable of giving useful partial information on molecular conformation. Each, of course, requires that the compound be actually available for study.

Molecular Orbital Prediction of Conformation

Another approach which has become available in the past decade is the use of all-valence electron, semiempirical molecular orbital theory. This approximation of quantum mechanics makes it possible to calculate for fairly large molecules, a total energy behaving in an approximately parallel fashion to the true molecular energy. The consideration of all valence electrons makes this calculated total energy sensitive to the conformation of the molecule. Thus, energy minimization as a function of bond angle variation is possible, and the prediction of a preferred conformation is a consequence.

The first of these methods was developed by Hoffmann in 1963 (*1*) and is known as extended Hückel theory (EHT). Briefly, the method uses Hückel formalism; however, explicit consideration of non-bonded interactions and all overlap integrals are a refinement. Slater orbitals are used, and the computations require only one parameter, the valence state ionization potential for the Coulomb integral and indirectly for the reso-

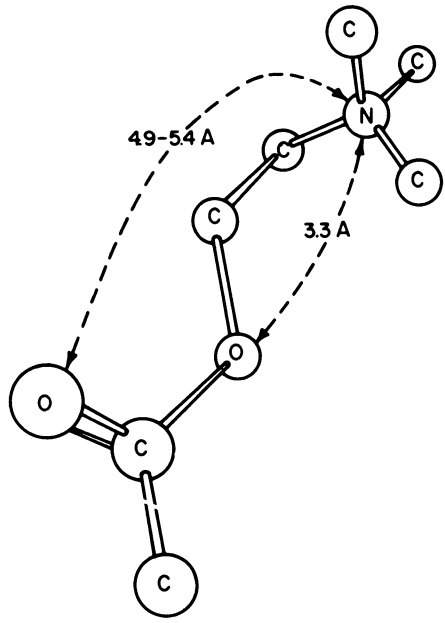


Figure 1. Predicted conformation of acetylcholine in "muscarinic" variant

nance integral. The theory, merits, and evaluations of EHT have been reported (*2–4*).

More recently, an all valence electron, semiempirical molecular orbital theory known as the Complete Neglect of Differential Overlap (CNDO) has been proposed by Pople based on self-consistent field (SCF) formalism (*5*). Although this method uses a more sophisticated approximation of the wavefunction, it neglects differential overlap.

Comparisons of these two methods reveal their relative strengths and shortcomings (*4*). In general, the CNDO method is superior for charges, EHT predicting greatly exaggerated values. The major value of EHT lies in its ability to predict correctly the preferred conformation. This has been demonstrated for numerous hydrocarbons (*1*) and more recently for a variety of heteroatom molecules (*6, 7*).

A large number of molecular orbital predictions of biological molecule conformations have been accomplished using EHT. The record of agreement between calculated and experimental values has been excellent. A significant amount of useful information has emerged from these predictions pertinent to the structure–activity relationships, the consideration of molecular mechanisms, and the rationale for new drug design (*7*). We attempt to summarize MO conformation studies through April 1972

and at the same time to assess the validity of these predictions in the light of experimental evidence and biological knowledge.

Muscarinic Agents

The first application of all-valence electron MO theory to predict the conformation of a neurotransmitter was reported in 1967 on acetylcholine (*8*). Using EHT, the conformation of acetylcholine was predicted to assume an approximately gauche relationship between the nitrogen and ether oxygen atoms. This is in agreement with experimental evidence derived from aqueous solution NMR studies (*9*). Some flexibility was predicted for the CO–O bond so that the dimensions between the heteroatoms were predicted as shown in Figure 1.

In the same study EHT predictions were reported for the potent muscarinic agents muscarine, Figure 2, and muscarone, Figure 3 (*8*). The experimental evidence available for comparison are x-ray analyses of the muscarine (*10*) and muscarone (*80*) crystals, with which the predicted conformations agree.

Superpositioning these three potent muscarinic agents, in their preferred conformations resulted in the observation of a common pattern of similarly charged structural features. The pattern, Figure 4, was proposed as the muscarinic pharmacophore (*8*). This pattern, predicted from theoretical considerations, bears a striking similarity to the muscarinic pharmacophore proposed by Beckett based on extensive structure–activity studies

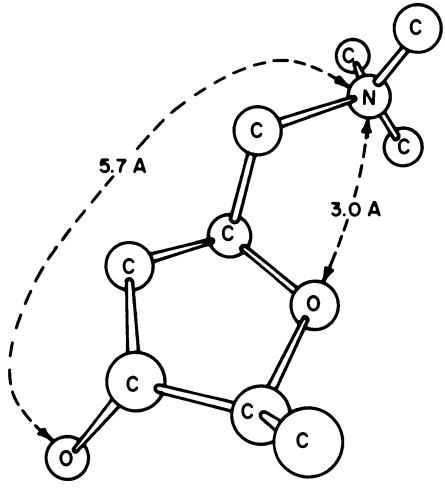


Figure 2. Predicted conformation of muscarine

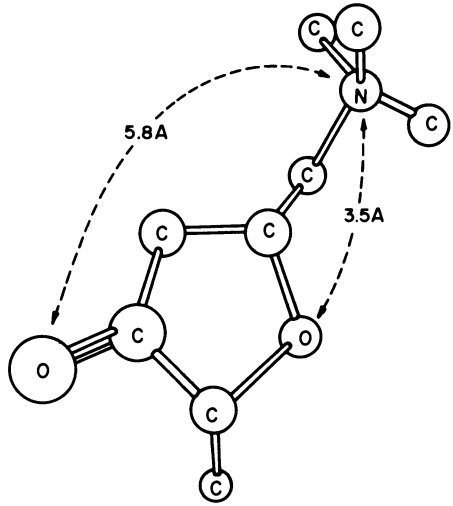


Figure 3. Predicted conformation of muscarone

(*11*). Recent MO calculations on these molecules using a modified CNDO method (*12*) and an intermediate Neglect of Differential Overlap (INDO) method (*13*) are in general agreement with EHT predictions and experimental evidence. A study of the potent muscarinic agent, S(+)-acetyl-β-methylcholine using both NMR and molecular rotation to predict the solution conformation, revealed that this molecule assumes the pattern predicted in Figure 4 (*14*), thus lending support to the validity of these predictions.

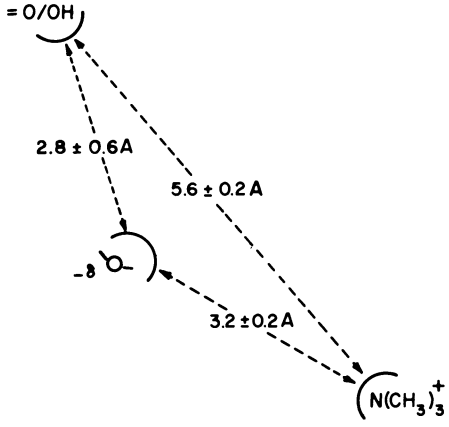


Figure 4. Proposed muscarinic pharmacaphore

Calculations on oxotremorine (*15*), presumed to be a CNS muscarinic agent, revealed that this molecule, in its predicted conformation (Figure 5) mimicks this predicted muscarinic pharmacophore. The assumption is made that the electronic character of the triple bond is the receptor equivalent of the ether oxygen atom in acetylcholine as proposed by Bebbington (*16*). These calculations support his proposal and reveal how oxotremorine may meet the structural requirements of a muscarinic agent.

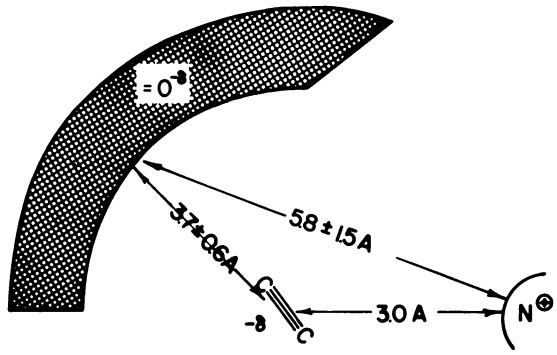


Figure 5. Predicted conformation of oxotremorine

Nicotinic Agents

The prediction of the preferred conformation of the nicotinic monocation was reported in 1968 (*17*). Two conformations were predicted to coexist (Figure 6) in agreement with NMR studies (*18*). Subsequent modified CNDO calculations are in agreement (*12*). One of these nicotine conformers, Figure 6 (top), presented an internitrogen distance close to the onium group–carbonyl oxygen distance previously predicted for acetylcholine. It is necessary, in this comparison, to invoke a predicted conformation of acetylcholine in which the flexible carbonyl group is rotated 60° as shown in Figure 7. It was predicted that Figure 6 (top) is the active conformer of nicotine and that Figure 7 is the nicotinic conformation of acetylcholine. The nicotinic pharmacophore was proposed to be as shown in Figure 8.

As a result of these two studies (*8*, *17*), it was proposed that acetylcholine exhibits two different pharmacological actions by virtue of its being capable of presenting to the receptors, two different groups of atoms in the acetylcholine molecule in two different conformations of the carbonyl group.

An EHT calculation on the potent nicotinic agent phenylcholine ether showed that this molecule prefers a gauche conformation of the hetero-

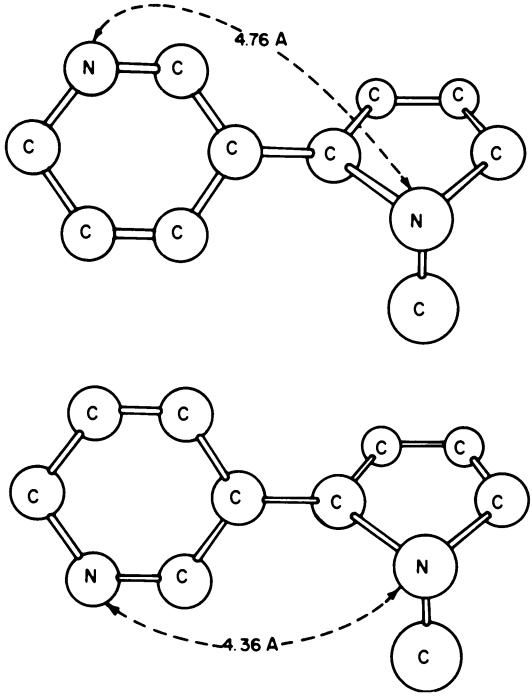


Figure 6. Predicted nicotine conformations

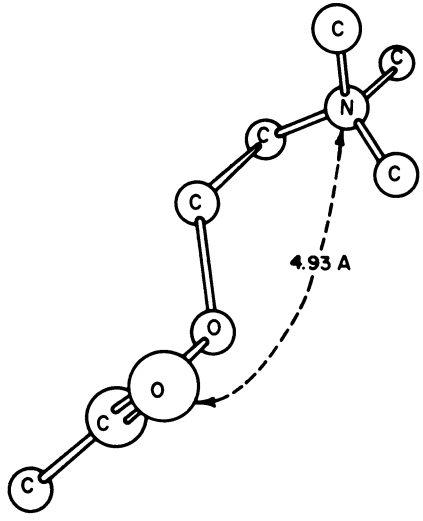


Figure 7. Predicted conformation of acetylcholine in "nicotinic" variant

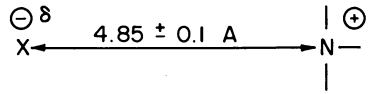


Figure 8. Proposed nicotinic pharmacophore

atoms, Figure 9 (*19*). The phenyl ring was found to prefer a conformation relative to the side chain so that an ortho position is about 4.7 A from the onium nitrogen. This prediction of a gauche side chain conformation in phenylcholine ether is in agreement with the crystal conformation found for xylocholine (*20*). Previous studies have shown a correlation between the frontier electron density (*21*) and the electrophilic reactivity (*22*) of this position and nicotinic activity. The conformational prediction of the molecule thus reveals that the ortho position is both spatially and electronically capable of fulfilling the structural requirements for nicotinic activity as previously proposed (Figure 8).

Recent EHT calculations on the neostigmine molecule have led to a conformation prediction locating the onium group about 4.5 A from the carbonyl oxygen (*81*) in support of the postulated nicotinic pharmacophore (*17*).

Cholinesterase Inhibitor

The first application of all-valence MO theory to a pharmacologically active agent was the EHT study of the cholinesterase inhibitor 2-formyl-*N*-methylpyridinium oxime (2-PAM$^+$) (*23*). The aldoxime group was found to prefer a conformation perpendicular to the ring plane. Crystallographic data indicate that the aldoxime group is coplaner with the ring in the solid state (*24*). The crystal is yellow while an acid solution is colorless, lending support to the belief that the MO calculations correctly predict a non-conjugated form and that calculations of this kind are relevant to dilute solution phenomena but not necessarily to crystal structure (*20*).

Histamine

Molecular orbital calculations led to the prediction that two distinctly different conformations of equal preference prevailed for histamine (Figure 10) (*25*). The conformation of Figure 10 (bottom) was predicted to exist without any hydrogen bonding between an onium hydrogen and the ring nitrogen atom. Recent NMR analysis of an aqueous solution of histamine revealed that the two conformers predicted from EHT–MO do in fact

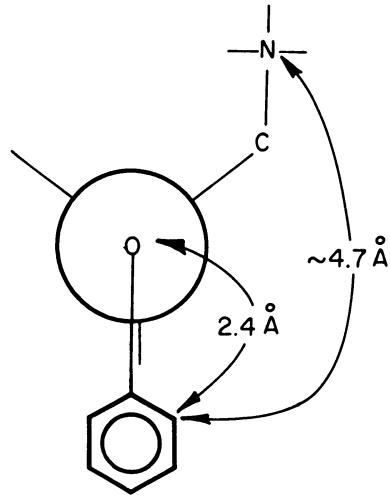


Figure 9. Predicted phenylcholine ether conformation

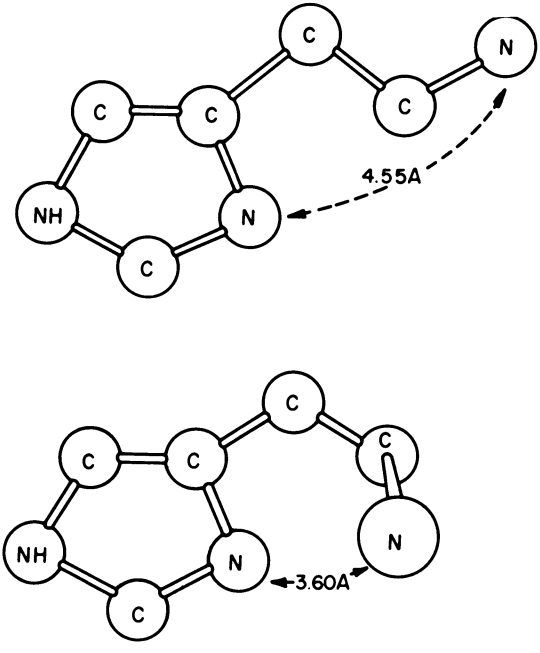


Figure 10. Predicted histamine conformations

exist in equal proportions (*26*). Furthermore, the prediction of no intramolecular hydrogen bonding in aqueous solution has also been confirmed experimentally (*27*). A recent modified CNDO calculation on histamine predicts only a single modified gauche conformation to predominate (*28*).

On the basis of the prediction of two coexisting conformers of histamine, the hypothesis was proposed that one conformer was responsible for the H_1-receptor activity and the other for H_2-receptor activity (*25*). The top conformer in Figure 10 was proposed as the H_1-receptor agonist on the basis of an internitrogen distance comparable with the presumed internitrogen distance in the antihistaminic triprolidine. The bottom conformer in Figure 10 was thus proposed as being the agonist for H_2-receptor activity. This latter prediction thus provides a rationale for the synthesis of potential acid antisecretory agents.

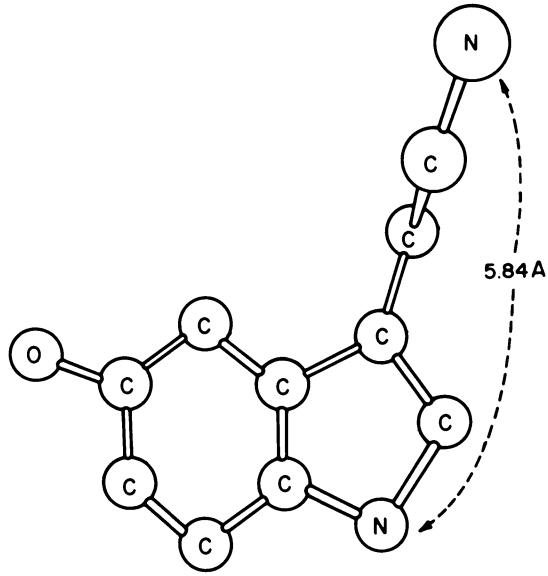


Figure 11. Predicted serotonin conformation

Serotonin

The preferred conformation of serotonin (5-hydroxytryptamine) was predicted using EHT–MO (Figure 11) (*29*). A modified CNDO calculation predicts a preferred conformation in which the C–N bond nearly eclipses the ring–C bond with the onium group lying over the benzene ring (*30*). An INDO calculation predicts a strong preference for a conformation in which the C–N bond eclipses the ring–C bond and the onium

group lies over the indole 2 position (*31*). A recent NMR analysis of the molecule (*82*) reports a trans preference, in support of the EHT study (*29*) and in contrast to the predictions by other methods. The EHT predicted distance between the two nitrogen atoms is very close to the internitrogen distance found in the competitive inhibitor lysergic acid diethylamide (LSD). It is reasonable to presume that a competitive inhibitor must partially mimic the pharmacophore of the agonist; hence these similar distances are consistent with the predicted conformation.

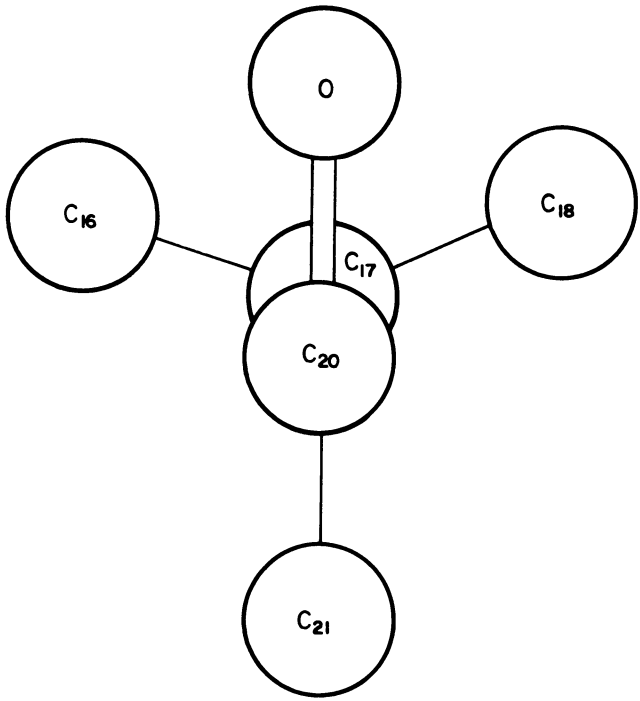


Figure 12. Predicted side chain conformations of 17-keto steroids

It is interesting to contemplate the structural similarity predicted for LSD and serotonin and the known CNS activity of the former and to compare this with the CNS activity and structure of the cannibinol metabolite, 11-hydroxy-Δ^9-tetrahydrocannabinol. It is conceivable that similar mechanisms may prevail in the CNS as a result of similar stereochemical presentations of comparable charged atoms to the serotonin receptor.

Pregnane Steroids

The side chain conformations of progesterone, corticosterone, and cortisol were predicted using EHT-MO (Figure 12) (*32*). All three side chains were predicted to form a plane with the 17α-substituent. The prediction of the progesterone side chain conformation is within 30° of a solution dipole moment study (*33*). The predictions of the corticosterone and cortisol conformations are in close agreement with experimental results from infrared and NMR studies (*34, 35*). A reflection of the β-face of cortisol, in its predicted conformation, is shown in Figure 13.

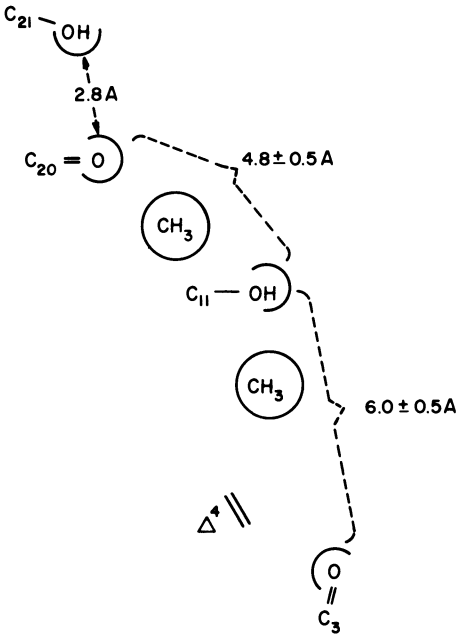


Figure 13. Predicted β-face pattern of cortisol

Portions of the predicted cortisol pattern of charged atoms (Figure 13) were observed to be comparable with charged patterns for either histamine (*25*) or serotonin (*29*). Since these two amines have been implicated as being inflammagenic (*36*) and cortisol is a potent anti-inflammatory agent (*37*), it was postulated that cortisol might evoke this action by an interaction with either or both histamine and serotonin receptors by virture of these common structural features.

The hypothesis has received some experimental support from the recent observation that cortisol is effective in competing for histamine binding sites on biopolymers (*38*). A subsequent EHT-MO study involv-

ing non-steroidal anti-inflammatory agents including indomethacin confirmed that the common structural patterns predicted prevailed in these drugs (*39*). The predicted conformation of indomethacin is in agreement with the reported crystal structure of this molecule (*83*). It is interesting to contemplate the roles of serotonin as a platelet aggregation promoter and the common anti-inflammatory agents as platelet aggregation inhibitors as a possible parallel to this hypothesis of anti-inflammatory activity.

Adrenergic Agents

EHT-MO predictions of ephedrine and pseudo-ephedrine have been reported (Figure 14) (*35*). The conformation predicted for ephedrine (*2*) is in agreement with NMR analysis (*41, 42*); however, the prediction of the pseudo-ephedrine conformation agrees only with a minor contributor to the solution equilibrium (*42*). The predicted conformation for the α-adrenergic agonist, norepinephrine, preser s the same pharmacophore (*43*). This prediction is in agreement with the crystal conformation (*44*). Other studies using the INDO (*45*) and CNDO (*84*) methods predict a trans and a gauche conformation to coexist.

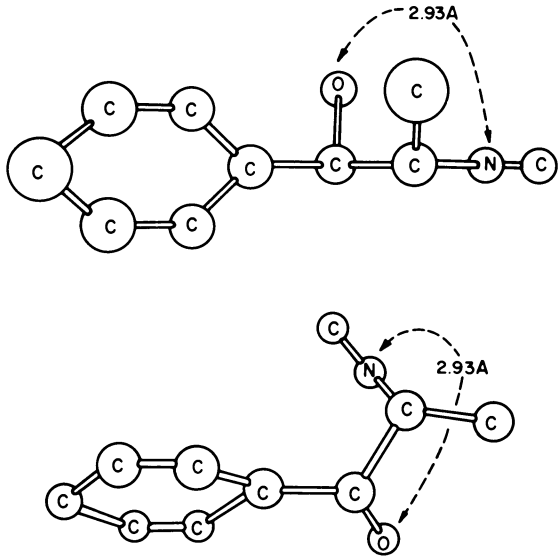


Figure 14. Predicted conformations of ephedrine and pseudoephedrine

It is well known that an unsubstituted catecholamine such as norepinephrine or a monomethyl derivative like epinephrine or ephedrine is predominantly α-adrenergic. Increasing the bulk of the mono alkyl sub-

stituent in this series increases β-adrenergic activity while at the same time α-adrenergic activity is obliterated. Thus, isopropyl norepinephrine (isoproteranol) is a standard for almost pure β-adrenergic activity.

Two theories have been proposed to explain this reversing pair of trends in the catecholamine series. One theory proposes that the increasing bulk of the N-substituent increases the barrier to rotation between the methylene groups so that the ease of assuming a gauche conformation will be a function of the N-substituent bulk (*46*). The theory then proposes that the different conformers, gauche and trans, cause two different reactions to occur, each characteristic of the α- or β-adrenergic receptor. The second theory proposes that the N-substituent influences the charge on the onium group which influences the reactivity to one or the other adrenergic receptor (*47*).

A recent study using EHT for conformation and CNDO and *ab initio* calculations for charge densities has been reported on the catecholamine series norepinephrine, epinephrine, *N*-ethylnorepinephrine, and isoproteranol (*48*). These studies revealed no predicted change in the trans preference for any of the series. Further, they showed an almost identical energy of the barrier from a trans to a gauche conformation for all members of the series. These results argue against the theory of variable flexibility of the methylene–methylene bond (*46*). Charge densities were calculated on simulated onium systems using both CNDO and *ab initio* methods (*48*). These results revealed no appreciable change in onium charge in the series, which is at variance with the charge theory (*47*).

An alternate theory was proposed, based on these calculations. It was postulated that α-adrenergic activity requires an onium hydrogen atom, probably as a hydrogen bond donor. The β-adrenergic receptor, however, was postulated to require an alkyl group at the N-substituent position. It was proposed that this N-substituent was involved in a dispersion interaction with the receptor and was optimal when there was a branched hydrocarbon such as an isopropyl group. These authors raised the intriguing suggestion that the onium group was perhaps not essential for β-adrenergic activity and that a methylene group could replace it. It was noted that such a compound had recently been made and reported (*49*). This compound, the methylene analog of isoproteranol was found to have modest β-adrenergic activity (*49*). A parallel was also drawn between this hypothesis and the structure and β-adrenergic activity of prostaglandin E_1 (*48*).

Dopamine

An EHT calculation on dopamine, considering the phenyl–methylene and methylene–methylene bonds and holding the hydroxyls out of the

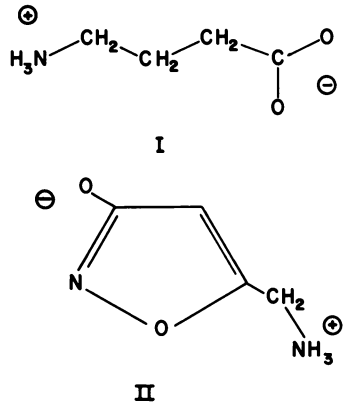


Figure 15. Predicted conformation of gamma-aminobutyric acid

ring plane led to a gauche prediction for the side chain (*50*). Another EHT calculation, apparently considering only the methylene–methylene bond, reported a trans preference (*85*). A CNDO-type calculation has reported a trans and gauche preference (*84*). Recalculation of the molecule holding the hydroxyls in the ring plane leads to a trans side chain conformation (*77*). Crystal analysis predicts trans (*86*) while NMR predicts both trans and gauche (*85*).

γ-Aminobutyric Acid

The preferred conformations of the inhibitory transmitter gamma-aminobutyric acid (GABA) and a GABA-like agent, muscimol, were calculated (*51*). The predicted conformations (Figure 15) reveal a close correspondence between the onium group and charged oxygen distance in the two molecules. It has recently been reported that bicuculline is a specific inhibitor of GABA (*52*). Examination of Dreiding models of bicuculline or the free acid reveal that this molecule, as a protonated salt, can assume a reasonable conformation in which the onium to oxygen distance is comparable with the prediction for GABA and muscimol. Recent calculations confirm this interatomic distance (*77*).

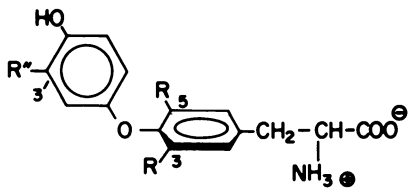
Thyromimetics

Trisubstituted thyronine analogs have been studied using EHT to predict the phenoxy phenyl conformation as influenced by the 3,5-substituents (Figure 16) (*53*). EHT predictions indicated that 3,5-iodo and 3,5-bromo groups influenced a preference as shown in Figure 16, in agree-

ment with crystal studies (*54, 87*). Recent reports that an iodine-free bromothyronine is thyromimetic suggest that the roles of these two halogens might be to contain the two rings as shown in Figure 16 (*55*). The MO calculations further predicted that 3,5-chloro or hydrogen analogs preferred alternate conformations to that shown in Figure 16, in which the two ring planes intersect like pages in a book. The knowledge that the chlorine or hydrogen analogs are inactive suggests a predominantly steric role for these atoms.

Amino Acids and Peptides

All valence electron MO calculations have been extended to the prediction of amino acid residue conformations. The approach has generally been to consider a model compound, such as an *N*-acylamino acid amide to simulate the mid-chain residue. Beginning with three independent studies reported in 1969 (*56–58*) a number of amino acid residue conformations have been predicted to date from all valence MO methods. Specific examples of amino acid residues studied, with references in order of appearance in the literature are glycine (*56–59*), alanine (*56, 57, 59*), phenylalanine (*57*), proline (*57, 60*), hydroxyproline (*60*), serine (*61, 62*), isoleucine (*61, 88*), valine (*61, 88*), threonine (*62*), leucine (*61, 88*), arginine (N-terminal) (*63*), arginine (C-terminal) (*63*), arginine side



a.

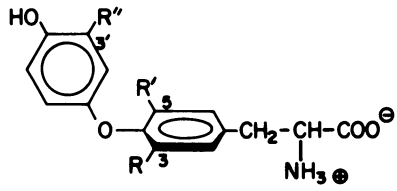


Figure 16. Predicted conformation of 3,5,3'-triiodo or tribromothyronine

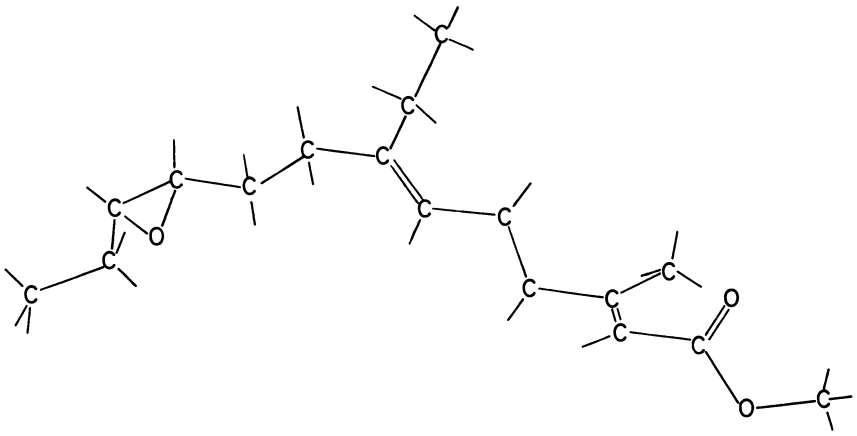


Figure 17. Predicted conformation of insect juvenile hormone

chain (*63, 64*), arginine and lysine residues (*65*), N-terminal glycine, alanine, and proline (*66, 89*), aspartic acid (*88*), glutamic acid (*88*), methionine (*90*), and tryptophan (*90*).

In general the predicted conformations have agreed with some experimental results obtained from various polypeptides using different experimental techniques. The existence of more than one form of a polypeptide resulting from different conformations possible for some amino acid residues, depending on physical and chemical manipulations, makes validation of these predictions complex.

Two attempts have been reported to predict the conformation of a polypeptide hormone by assembling the appropriate residues in their predicted conformation (*63, 90*). In this manner, the amino acid sequence of bradykinin was predicted to exist in a random coil conformation, with variation around the glycine ϕ bond, and with no interaction predicted between phenyl groups. As yet no experimental evidence confirms this prediction; however, existing experimental evidence suggests that the prediction is reasonable (*67–69*). In the second study the conformation of gastrin tetrapeptide was predicted (*90*).

Other Biologically Important Molecules

A number of other studies have been reported, using all-valence MO methods to predict the conformation of biologically important molecules. These include calculations on glucopyranose (*70*), several disaccharides (*71*), several nucleosides (*72–75*), and acetanilide (*76*). The prediction of the conformation of the insect juvenile hormone has also been made from EHT-MO calculations (Figure 17) (*77*). Using a combination of

EHT, iterative EHT, and dispersion bonding calculations, the conformations of prostaglandin E-1 were predicted (*78*). The prominent conformers were all predicted to have intimate interaction between the side chains, in agreement with crystal studies (*79*). Finally, predictions of conformation have led to the prediction of a sweet-taste pharmacophore (*91*).

Perspectives

The relevance of the preferred conformation of an active molecule to events at its biological receptor is uncertain. The finding in several studies, that structurally different molecules with equal potencies prefer conformations in which very similar patterns of charged atoms are made available, lends support to the relevance. It has been suggested that the drug molecule forms a preliminary weak bond with the receptor while in the preferred conformation, followed by a more intimate and stronger association in which both drug and receptor conformations are altered (*50*). The stereoselective "recognition" of the drug, however, occurs during the preliminary event, prior to any mutual perturbation. Much remains to be explored concerning these events; however, results to date suggest that the presumption of the importance of preferred conformation is a useful operational hypothesis.

Finally, MO predictions should be considered as an adjunct to the drug design and mechanism interpretation process. The calculations should always be based on reasonable chemical structures and processes, and the results and interpretations should always be viewed in the light of biological reality or reasonable mechanisms. The MO theory can thus be a powerful servant to the medicinal chemist or chemical pharmacologist in his search for biological explanations and new drugs.

Literature Cited

1. Hoffmann, R., *J. Chem. Phys.* (1963) **39**, 1397.
2. Allen, L. C., Russell, J. D., *J. Chem. Phys.* (1967) **46**, 1029.
3. Blyholder, G., Coulson, C. A., *Theoret. Chim. Acta* (1968) **40**, 316.
4. Hoyland, J. R., in "Molecular Orbital Studies in Chemical Pharmacology," L. Kier, Ed., Springer-Verlag, New York, 1970.
5. Pople, J. A., Santry, D. P., Segal, G. A., *J. Chem. Phys.* (1965) **43**, 5129.
6. Kier, L. B., in "Fundamental Concepts in Drug-Receptor Interactions," J. F. Danielli, J. F. Moran, D. J. Triggle, Eds., Academic Press, New York, 1970.
7. Kier, L. B., "Molecular Orbital Theory in Drug Research," Academic Press, New York, 1971.
8. Kier, L. B., *Mol. Pharmacol.* (1967) **3**, 487.
9. Culvenor, C., Ham, N., *Chem. Commun.* (1966) 537.
10. Jellinek, F., *Acta Cryst.* (1957) **10**, 277.

11. Beckett, A. H., Harper, N. J., Clitherow, J. W., *J. Pharm. Pharmacol.* (1963) **15,** 362.
12. Pullman, B., Courriere, P., Coubeils, J. L., *Mol. Pharmacol.* (1971) **7,** 397.
13. Beveridge, D. L., Radna, R. J., *J. Am. Chem. Soc.* (1971) **93,** 3759.
14. Inch, T. D., Chittenden, R. A., Dean, C., *J. Pharm. Pharmacol.* (1970) **22,** 954.
15. Kier, L. B., *J. Pharm. Sci.* (1970) **59,** 112.
16. Bebbington, A., Brimblecomb, R. W., Shakeshaft, D., *Advan. Drug Res.* (1965) **2,** 143.
17. Kier, L. B., *Mol. Pharmacol.* (1968) **4,** 70.
18. Simpson, T. R., Craig, J. C., Kumler, W. D., *J. Pharm. Sci.* (1967) **56,** 708.
19. Kier, L. B., *J. Med. Chem.* (1971) **14,** 80.
20. Coggon, P., McPhail, A. T., Roe, A. M., *Nature* (1969) **224,** 1200.
21. Fukui, K., Nagata, C., Imamura, A., *Science* (1960) **132,** 87.
22. Crow, J., Wasserman, O., Holand, W. C., *J. Med. Chem.* (1969) **12,** 764.
23. Giordano, W., Hamann, J., Harkins, J., Kaufman, J., *Mol. Pharmacol.* (1967) **3,** 307.
24. Carlstrom, D., *Acta Chem. Scand.* (1966) **20,** 1240.
25. Kier, L. B., *J. Med. Chem.* (1968) **11,** 441.
26. Casy, A. F., Ison, R. R., Ham, N. S., *Chem. Commun.* (1970) 1343.
27. Paiva, T. B., Tominaga, M., Paiva, A. C. M., *J. Med. Chem.* (1970) **13,** 689.
28. Coubeils, J. L., Courriere, P., Pullman, B., *Compte Rend. Acad. Sci., Paris* (1971) **272,** 1813.
29. Kier, L. B., *J. Pharm. Sci.* (1968) **57,** 1188.
30. Courriere, P., Coubeils, J. L., Pullman, B., *Compte Rend. Sci., Paris* (1971) **272,** 1697.
31. Kang, S., Cho, M. H., *Theoret. Chim. Acta* (1971) **22,** 176.
32. Kier, L. B., *J. Med. Chem.* (1968) **11,** 915.
33. Allinger, N. L., DaRooge, M. A., *J. Am. Chem. Soc.* (1961) **83,** 4256.
34. Cole, W. G., Williams, D. H., *J. Chem. Soc.* (1968) 1849.
35. Cole, W. G., Williams, D. H., *J. Chem. Soc.* (1970) 748.
36. Spector, W. G., Willoughby, D. A., *Ann. N. Y. Acad. Sci.* (1964) **116,** 839.
37. Winter, C. A., *Progr. Drug Res.* (1966) **10,** 139.
38. Botre, C., Marchetti, M., Del Vecchio, C., Lionetti, G., Memoli, A., *J. Med. Chem.* (1969) **12,** 832.
39. Kier, L. B., Whitehouse, M. W., *J. Pharm. Pharmacol.* (1968) **20,** 793.
40. Kier, L. B., *J. Pharmacol. Exptl. Therap.* (1968) **164,** 75.
41. Hyne, J. B., *Can. J. Chem.* (1961) **39,** 2536.
42. Portoghese, P., *J. Med. Chem.* (1961) **39,** 2536.
43. Kier, L. B., *J. Pharm. Pharmacol.* (1969) **21,** 93.
44. Carlstrom, D., Bergin, R., *Acta Cryst.* (1967) **23,** 313.
45. Pedersen, L., Hoskins, R. E., Cable, H., *J. Pharm. Pharmacol.* (1971) **23,** 216.
46. Larsen, A. A., *Nature* (1969) **224,** 25.
47. Pratesi, P., Grava, E., *Adv. Drug Res.* (1965) **2,** 127.
48. George, J. M., Kier, L. B., Hoyland, J. R., *Mol. Pharmacol.* (1971) **7,** 328.
49. Villa, L., Ferri, V., Grana, E., Mastelli, O., *Il Farmaco, Ed. Sci.* (1970) **25,** 118.
50. Kier, L. B., Truitt, E. B., *J. Pharmacol. Exptl. Therap.* (1970) **174,** 94.
51. Kier, L. B., Truitt, E. B., *Experientia* (1970) **26,** 988.
52. Curtis, D. R., Duggan, A. W., Felix, D., Johnston, G. A. R., *Nature* (1970) **226,** 1222.
53. Kier, L. B., Hoyland, J. R., *J. Med. Chem.* (1970) **13,** 1182.
54. Cody, V., Duax, W. L., Norton, D. A., *Chem. Commun.* (1971) 683.
55. Jorgensen, E. C., Nulu, J. R., *J. Pharm. Sci.* (1969) **58,** 1139.
56. Hoffmann, R., Imamura, A., *Biopolymers* (1969) **7,** 207.
57. Kier, L. B., George, J. M., *Theoret. Chim. Acta* (1969) **14,** 258.

58. Rossi, A., David, C. W., Schor, R., *Theoret. Chim. Acta* (1969) **14,** 429.
59. Maigret, B., Pullman, B., Dreyfus, M., *J. Theor. Biol.* (1970) **26,** 321.
60. Maigret, B., Perahia, D., Pullman, B., *J. Theor. Biol.* (1970) **29,** 275.
61. George, J. M., Kier, L. B., *Experientia* (1970) **26,** 952.
62. Perahia, D., Maigret, B., Pullman, B., *Theoret. Chim. Acta* (1970) **19,** 121.
63. Kier, L. B., George, J. M., in "Molecular Orbital Studies in Chemical Pharmacology," L. B. Kier, Ed., Springer-Verlag, New York, 1970.
64. Ponnuswamy, P. K., Lakshminarayanan, A. V., Sasisekharan, V., *Biochim. Biophys. Acta* (1971) **229,** 596.
65. Pullman, B., Coubeils, J. L., Courriere, P., Perahia, D., *Theoret. Chim. Acta* (1971) **22.**
66. Bossa, M., Damiani, A., Fidenzi, R., Gigli, S., Leli, R., Ramunni, G., *Theoret. Chim. Acta* (1971) **20,** 299.
67. Brady, A. H., Stewart, J. M., Ryan, J. W., in "Cardiovascular and Neuroactions of the Kinins," N. Back, F. Sicuteri, M. Rocha de Silva, Eds., Plenum Press, New York, 1970.
68. Bodanszky, A., Bedanszky, M., Torpes, E. J., Mutt, V., Ondette, M. A., *Experientia* (1970) **26,** 948.
69. Brady, A. H., Ryan, J. W., Stewart, J. M., *Biochem. J.* (1971) **121,** 179.
70. Neely, W. B., *J. Med. Chem.* (1969) **12,** 16.
71. Giacomini, M., Pullman, B., Maigret, B., *Theoret. Chim. Acta* (1970) **19,** 347.
72. Jordan, F., Pullman, B., *Theoret. Chim. Acta* (1968) **9,** 242.
73. Cameran, N., Nyburg, S. C., *Theoret. Chim. Acta* (1969) **13,** 162.
74. Berthod, H., Pullman, B., *Biochim. Biophys. Acta* (1971) **232,** 595.
75. Govil, G., Saran, A., *J. Theor. Biol.* (1971) **30,** 621.
76. Olsen, J. F., Kang, S., *Theoret. Chim. Acta* (1970) **17,** 329.
77. Kier, L. B., unpublished data.
78. Hoyland, J. R., Kier, L. B., *J. Med. Chem.* (January 1972).
79. Rabinowitz, I., Ramwell, P., Davison, P., *Nature New Biol.* (1971) **223,** 88.
80. Pauling, P., Petcher, T. J., *Nature New Biol.* (1972) **236,** 112.
81. Crow, J. W., Holland, W. C., *J. Med. Chem.* (1972) **15,** 429.
82. Ison, R. R., Partington, P., Roberts, G. C. K., *J. Pharm. Pharmacol.* (1972) **24,** 84.
83. Kistenmacher, T. J., Marsh, R. E., *J. Amer. Chem. Soc.* (1972) **94,** 1340.
84. Pullman, B., Coubeils, J. L., Courriere, P., Gervois, J. P., *J. Med. Chem.* (1972) **15,** 17.
85. Bustard, T. M., Egan, R. S., *Tetrahedron* (1971) **27,** 4457.
86. Bergin, R., Carlstrom, D., *Acta Cryst.* (1968) **B24,** 1506.
87. Camerman, N., Camerman, A., *Science* (1972) **175,** 764.
88. Maigret, B., Perahia, D., Pullman, B., *Biopolymers* (1971) **10,** 491.
89. Bossa, M., Damiani, A., Fidenzi, R., Gigli, S., Leli, R., Ramunni, G., *Theoret. Chim. Acta* (1971) **20,** 299.
90. Kier L. B., George, J. M., *J. Med. Chem.* (1972) **15,** 384.
91. Kier, L. B., *J. Pharm. Sci.,* in press.

RECEIVED May 10, 1971.

INDEX

D

E

F

G

Q

R

S