Comparative QSAR: Toward a Deeper Understanding of Chemicobiological Interactions

Corwin Hansch,* David Hoekman, and Hua Gao

Department of Chemistry, Pomona College, Claremont, California 91711

Received August 30, 1995 (Revised Manuscript Received January 24, 1996)

Contents

I. Introduction

A major problem, which has become acute in the latter part of the 20th century in chemistry and, indeed, in all areas of intellectual endeavor, is the "information explosion". How much can one mind attempt to encompass? How many books out of the millions published (and, most importantly, which ones) can one read in a lifetime? In 1994 Chemical Abstracts published 653 055 abstracts—1789 per day! Miss ten days and you are behind 17 890 articles or patents. The problem is not so much getting information, but how do we organize it? How do we make sense of it? So too with how chemicals react with themselves and in turn how do these reactions affect organisms or parts thereof: DNA, enzymes, organelles, etc.

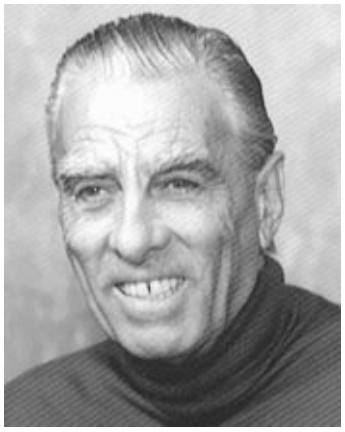
Many databases have been built listing drugs, pesticides, carcinogenic and mutagenic chemicals, skin and eye irritants, environmental toxicants, etc. We are lost in a sea of data. This is not meant to disparage such efforts since they are the first step to outlining the problem. The next step is to formulate equations rationalizing small regions of structure–activity space. The success of the Hammett equation, in this respect, has been truly remarkable. The vast number of such relationships provides a cornerstone on which to relate biological quantitative structure–activity relationships (QSARs) to the better understood areas of physical organic chemistry.

In this review we consider a new approach to reviewing the literature on quantitative structure–activity relationships (QSARs) and *apply it only to biological QSAR which contain a term in* σ *or* σ^-. Traditional methods of review, such as those published in this journal, do an excellent job on the various subsections of our science; however, we have reached the point where broader generalizations can be experimented with. One such area is that of QSAR. Many thousands have been published in hundreds of different journals ranging from chemical physics to psychobiology and environmental toxicology. Simply collecting and storing the data is not enough. How does one review what has been captured in, say, 15 000 equations based on several hundred thousand data points? This is especially important in the design of bioactive compounds and toxicology. While the convenient storage and retrieval of the activity of the individual compounds is, of course, vital, the predictive ability of equations is needed to further the construction of a science of QSAR. Considering the almost infinite number and variety of organic chemicals means that only an infinitesimally small percent of them can be tested in an extremely limited set of biological tests to establish their activity. For instance, if one used 166 substituents to substitute the 7 positions on the quinoline ring this would yield approximately 10^{15} "congeners". Of course 166 is a very small fraction of the number of possible substituents. We could easily design 1000. The number of organisms, enzymes, cells, organelles, etc., on which these could be tested is also more or less endless. To begin to organize our thinking on such a staggering problem we must effectively use what has been published. What has been learned in one area must be related to other areas. We can ill afford redundancy in either the testing or synthesis of organic chemicals.

Structure–activity relationships have been at the heart of chemistry since Mendeleyev began the construction of the periodic table. It was Hammett around 1935, and then his countless successors, who showed how equations could be derived to quantitatively describe and predict how organic compounds would undergo a given reaction. It is the extension of Hammett's thinking that forms the basis of our review.[1] Often it is possible to relate the biological QSAR to Hammett equations from physical organic chemistry. Bearing in mind that much drug research and all pesticide research is based on selective toxicity (toxicity to the pathogen but not to man) most QSARs center on some kind of toxic or inhibitory action.

At present two broad approaches to QSAR are evolving. One is the "traditional" method using experimental parameters for substituent effects; that is, Hammett constants (and sometime MO parameters) for electronic effects,[1] molar refraction, molar volume, or Verloop et al.'s sterimol parameters (or E_S) for steric effects, topological indices, and partition coefficients first used by Meyer and Overton for hydrophobic interactions.[2] The alternative approaches are with parameters which provide one with little insight in terms of chemical thinking as we know it.

0009-2665/96/0796-1045$25.00/0



Corwin Hansch received his undergraduate education in chemistry at the University of Illinois and his Ph.D. in organic chemistry from New York University in 1944. After working with the DuPont Company, first on the Manhattan Project and then in Wilmington, DE, he joined the Pomona College faculty in 1946. He has remained at Pomona College except for two sabbaticals—one at the Federal Institute of Technology in Zurich and the other at the University of Munich. His primary interest is in the relationship between the structure of organic compounds and their biological properties.

David Hoekman was born in 1961 in Holland, MI. He attended Pomona College for five years starting in 1980 and initially majored in physics. After three years of study he switched to biology and received his B.S. in biology in 1985. He spent a year working on ecological wood anatomy with Sherwin Carlquist at Rancho Santa Ana Botanic Garden. After a year of study in the botany department at the University of California at Berkeley, he turned his attentions toward computer programming. Since 1987 he has been the head of computer operations at the MedChem Project, where he designed and implemented a QSAR database and analysis package.

One of the difficulties with the classical parameters including those obtained from molecular orbital calculations is that they are not normally orthogonal. The degree of collinearity varies considerably depending on how they have been selected and the number used to formulate an equation, but in a general sense there is no way to completely eliminate collinearity. One can conclude that we have not yet been able to formulate true electronic, steric, or hydrophobic parameters; or, actually, that these three terms have no fundamental meaning as they are now used. They are parameters of the moment that will eventually give way to the truly fundamental parameters needed to describe chemical and chemicobiological reactions.

In an alternative approach to QSAR, Wold and his associates have solved the collinearity problem using the PLS (partial least squares) methodology.[3] They

Hua Gao was born in 1961 in Shandong, China. He studied Pharmaceutical Sciences at Shandong Medical University of China where he received his B.S. and M.S. degrees in Pharmaceutical Sciences. He joined Dr. Eric J. Lien of University of Southern California as a research scholar supported by Boeringer Ingelheim Funds of Germany in 1989. In 1995, he received his Ph.D. from University of Southern California in the field of Pharmaceutical Sciences. Currently, he is a postdoctoral researcher with Professor Corwin Hansch. His research interests include QSAR, computer-assisted drug design, and natural products.

use the largest possible set of descriptors of all kinds: melting point, molecular weight, partition coefficient, bond length and angles, etc., as well as molecular orbital parameters for a set of bioactive compounds. By extracting the principal components of the data matrix one solves the collinearity problem and hopefully obtains a set of descriptors which more completely covers the physicochemical characteristics of the compounds. This can result in many principal components, most of which are insignificant. The important ones, called latent variables, are selected by a process of common sense and regression analysis.

The price one pays for this approach is that the vectors so derived bear no relation to chemistry as it has developed to this point in time. The results cannot be related to mechanistic physical organic or biochemistry as these subjects are now understood. Hence it is virtually impossible to compare QSARs from different sets of chemicals acting on different or even similar biological systems, or to compare results from biological systems with those from organic reactions. In addition, it is very difficult to decide what should be the next derivative to test to obtain increased activity. Latent variables vary from system to system so that it is not possible to think in such consistent terms as it is with Hammett or MO parameters or π or log P.

The development of comparative molecular field analysis (CoMFA) by Cramer and his colleagues[4] also depends heavily on PLS. In the CoMFA approach all parameters are normally calculated, although log P is sometimes introduced. Hundreds or thousands of parameters are calculated to correlate 10−50 compounds. These must be reduced by the PLS method to the few significant latent variables.

There are many other approaches to QSAR[5] such as distance geometry, neural networks, and connectivity, to mention a few, which are not easy to relate from equation to equation. In fact there are so many different ways to obtain a statistically significant QSAR that even those specializing in the

subject are often confused as to the value of any particular equation. Unless we can show that there are meaningful relationships among QSARs for different chemicals acting on the same or different systems we cannot call it a science. Establishing lateral correlations among QSARs is the only path to developing such a science. Most of the thousands of published QSARs are based on the traditional parameters. Although our subject is bound to evolve in many ways unforseen, the order that we can begin to establish now will guide its development. For example, while it is unlikely that octanol−water partition coefficients are the ultimate means for defining hydrophobic interactions (indeed the term hydrophobic will eventually be better defined) the last 30 years of work has clearly demonstrated its empirical value. We now look askance at an equation which contains a coefficient (h) with a π or log P term greater than 1.2.[6] Of course this holds only for the linear or the initial slope in bilinear equations.[7] This is what we expect from observing hundreds of published and unpublished QSARs. If h falls outside this limit one suspects that something more than the usual hydrophobic effect is involved.

Although we started to organize QSARs with a view to making comparative studies some time ago,[8,9] it is only with the development of highly effective interactive computing and Internet for the exchange of data that a serious attempt can be made to organize what has been learned about QSAR since Hammett's initiation of the use of σ constants in 1935. The only attempt to list all Hammett equations was made in 1953 by Jaffe[10] who organized some 400 examples. At present our database contains over 4000 equations from physical organic chemistry, and we estimate that 2 or 3 times that number have been published or could be derived from published data. Our database of biological QSAR contains over 3000 examples which could also be increased by a factor of 2 or 3. It is impossible to make proper use of this vast amount of information without an interactive computerized system. It is already clear that with such a system we can begin to establish relationships between mechanistic physical organic chemistry and biological QSARs which do much to enhance our understanding of the latter.[1,11−13]

II. Structure of the System

In designing for the review and comparison of mathematical correlations of every type of organic compound (and even some inorganic) with every type of biological system there are a number of factors which need to be considered. Searching from every significant point of view is foremost. Labels must be standardized insofar as possible, but new labels must be readily accommodated.

A problem which will become more serious as the database increases in size is that of analyzing the results from any given search. That is, one generally wants to examine the smallest output necessary to uncover and establish a particular point. Having to consider hundreds or eventually thousands of hits can be time consuming, although sometimes there may be no way to avoid it. As we show below, focused searching can be time saving.

Often, interest is in data from physical organic *or* biological systems, hence our databank is divided into these two classes. This saves searching time, but when needed the two sets can be combined and searched simultaneously. Table 1 shows how each of the two major divisions has been divided for more specific searching. For example, searching on B3 sequesters all data on organelles. If only chloroplast data was desired, it could be isolated by searching on B3C. A number of subsets can be isolated in one search. It might be interesting to compare what has been done on oxidoreductases, mitochondria and animals, and this could be accomplished by isolating the QSARs from B2A, B3A, and B6A. Comparisons can be made between the two major divisions. It could be interesting to compare equations on microsomes (B3B) with those from oxidation (P10) or radical reactions (P12). Once such groups of QSARs have been isolated, they can be organized with respect to the terms in the equations or the coefficients with these terms.

The information associated with each dataset is outlined in Table 2. For biological data we have not attempted to standardize names in field 1 (system). This is accomplished with field 2 (class) in Table 2. One or more classes from Table 1 are assigned to each set. For instance, for a QSAR on curing mice of an *Escherichia coli* infection one could assign classes B4B and B6A to the set. Under the heading system a common name such as mouse, fly, heart, and serum albumin has been entered. Eventually it may be desirable to bring more order to field 1. In field 3 we have assigned common chemical names or a linear structure, i.e. benzodiazepines or X-C_6H_4-COOR. The precise 2-D structure is given in field 12 via the SMILES notation.[14] Again in the case of action (field 4) standardization has not yet been attempted, instead terms such as LD_{50}, antiinflammatory, diuretic, antitumor, antimalarial, etc., have been used. More chemically related names such as demethylation, hydrolysis, oxidation have been used when possible. A browsing mode can be used to inspect the names in any field to get an idea of what is in the system. In field 5 a standard reference to the literature source of the data set is supplied. Field 8 (note) is especially important since it is here that one must define any special terms in a QSAR or provide any information which might be of help in understanding the test system. Once a search has been made any or all of the data of Table 2 can be displayed for each hit. That is, after a search has been completed one goes to the "show" mode where it is displayed always with an assigned set number. This number is used to retrieve the set in a regression mode where it can be restudied with new parameters at any time.

The physical organic data has been organized in much the same way except that in field 1 (system) the solvent is given. Under action, descriptors such as hydrolysis, oxidation, Beckman rearrangement, and Friedel−Crafts reaction are used.

III. Database Searching

String searching is generally used. For example searching field 1 for **acetylcholinesterase** makes

Table 1. Biological and Physical Organic Classifications

	Biological		
B1	**nonenzymatic macromolecules** (DNA, fibrin, hemoglobin, albumin)	**B4**	single-celled organisms
B1P	plasma	B4E	erythrocytes
B2	**enzymes**	B4F	fungi, molds
B2A	oxidoreductases	B4L	leukocytes
B2B	transferases	B4P	protozoa
B2C	hydrolases	B4V	Viruses
B2D	lyases	B4Y	Yeasts
B2E	isomerases	**B5**	**organs/tissues**
B2F	ligase	B5C	Cancer
B2G	receptors	B5G	Gastrointestinal tract
B3	**organelles**	B5H	Heart
B3A	mitochondria	B5I	Internal/soft Organs
B3B	microsomes	B5L	Liver
B3C	chloroplasts	B5N	Nerves, Brain, Muscles
B3M	membranes	B5S	Skin
B3R	ribosomes	**B6**	**multicellular organisms**
B3S	synaptosomes	B6A	Animal (vertebrates)
B4	**single-celled organisms**	B6B	Insect
B4A	algae	B6F	Fish
B4B	bacteria	B6H	Human
B4C	cells in culture	B6I	Invertebrates
		B6P	Plant

	Physical Organic		
P0	**unknown**	**P7**	**addition**
P1	ionization	P7N	nucleophilic addition
P1P	ionization potential	P7P	polymerization
P1X	proton exchange	**P8**	**elimination**
P2	**hydrolysis**	**P9**	**rearrangement**
P3	**solvolysis**	**P10**	**oxidation**
P4	**spectra**	**P12**	**Radical Reactions**
P4I	ionization spectra	**P13**	**Complex Formation**
P4E	ESR spectra	**P14**	**Partitioning**
P4M	mass spectra	P14C	Chromatography
P4N	NMR spectra	**P15**	**pyrolysis**
P4R	IR spectra	**P16**	**H-bonding**
P4U	UV spectra	**P17**	**electrochemical**
P5	**miscellaneous reactions**	**P18**	**Brønsted**
P6	**substitution**	**P19**	**Esterification**
P6E	electrophilic substitution	**P20**	**Photochemical**
P6N	nucleophilic substitution	**P21**	**Hydrogenation**
P7	**addition**	**P22**	**Isokinetic**
P7D	dimerization	**P23**	**Reduction**
P7E	electrophilic addition		

12 hits. However, searching on **cholinesterase** identifies 91 sets while searching on **choline** finds 94 examples. That is, every word containing the string of letters in choline is identified. This can be very helpful in searching field 5 for an author's name. Often one cannot remember the exact spelling. String searching on **han** finds: Hansen, Hanson, Hansch, Khan, Shanke, etc. Searching by author is extremely helpful; so often one can remember the name of an author, but not where or when the paper was published.

Reference can also be searched by year: **5 (1990) (1991) (1992) (1993)** finds all references from the years 1990 through 1993. For instances where an author has published many papers, the list can be shortened by including the year and/or a second author's name.

To circumvent the string search one encloses the word or a fraction of it with quotes: **"choline"**. Either the first or second " can be omitted to allow string searching on the beginning or end of the word. Although all of the data from Table 2 can be displayed for any hit, one is rarely interested in fields 6−14 and 20. An important feature of the program is that any single field can be displayed. This becomes very important when hundreds (eventually thousands) of sets of interest are found. Often by simply displaying one field, say system or compound, one can recognize sets of interest. One line associated with each set can be perused for hundreds of sets in minutes, whereas looking at fields 1−5 and 15, 16, and 18 could take hours.

There are three major ways to search the database:

(1) The most direct way to narrow one's focus is to search on one or a few of the labels of Table 1. The number of possibilities is large. For instance comparing the 36 subsets of the biological section of Table 1 two at a time would mean 630 different comparisons. This number would be greatly increased by including the subsets of the physical organic database.

(2) QSARs can be sequestered according to the parameters they contain and then compared by means of the categories of Table 1 or 2. For instance, the combined physical organic and biological banks could be searched for all QSARs containing a term in σ by searching on **" S "**. At present this finds almost 2000 sets in the physical organic database (note that this method of searching would not hit equations containing σ^-, σ^+, or σ_I). If necessary these could be

Table 2. Organization

field	title	description
		Data Set
1	SYSTEM	biological or physical organic system
2	CLASS	see Table 1
3	COMPOUND	parent compound
4	ACTION	measured action or activity
5	REFERENCE	journal reference or other source of data set
6	SOURCE	person who entered data set
7	CHECK	person who checked data set
8	NOTE	additional information about data set
9	DATE	date on which set was saved into database
10	PARAMETERS	list of parameters
11	SUBSTITUENTS	list of substituents or structures
12	SMILES	topological description of compounds
13	DATA	table of parameter values
14	PRM MAX/MIN	maximum and minimum of each parameter
		Equation
15	TERMS IN EQN	parameters in regression equation
16	EQUATION	regression coefficients for each parameter
17	IDEAL	ideal, log *P*, or other parameters and confidence limits
18	STATISTICS	*n*, d*f*, *r*, *s*
19	RESIDUALS	deviations between *y* predicted and *y* observed
20	PREDICTED	predicted values of dependent parameter

examined in the show mode by sorting on the coefficient with $\sigma(\rho)$ and displaying the results in order of increasing value of ρ. Asking to see only one or two fields of Table 2 (such as terms of equation and action or compound or system) one could in a reasonable time consider all sets. If one has some idea of the size of ρ of interest then range searching can be employed. For example we could isolate all sets containing ρ in the range -1.0 to $+2.0$ or even a narrow range of 2.0 to 2.1 and these could then be displayed in order of increasing size of ρ. This technique could be used for any parameter such as log *P*, σ^-, E_s, etc. Actually one might isolate equations containing two or more parameters and then display them in order of increasing coefficient size on any one parameter.

(3) The third general mode for searching and comparing is based on chemical structure. Two different avenues are open. By using field 12 a search can be made for any particular compound. For example searching the biological database on the SMILES for phenol makes 162 hits. This means that 162 sets contain phenol, but the set may be composed of miscellaneous chemicals. These can be eliminated by searching on **3 not misc** which reduces the number of sets to 127. Now in the show mode displaying via **3** we still find sets with other compounds (e.g. benzenes and phenols) searching on **3 not benzene** yields a group of 117 sets. This could be narrowed further, for example by means of field 18. The search command **18 n>15** locates all sets with more than 15 compounds.

Another important means for searching involves the use of the **not** command. For instance, searching with **15 not logP Pi RM** on the biological database finds all equations (and associated data) which do not contain hydrophobic terms (RM is the chromatographic means of defining hydrophobicity). At present there are 470 examples. String searching catches terms such as ClogP, logP′, Pi-sum, etc., and these sets are eliminated (ClogP represents calculated log *P* values).

While the above searching mode identifies sets in which phenol is present, a second method of substructure searching also based in SMILES identifies many individual compounds. Again the structure of phenol is used, but now every aromatic structure containing a phenolic $-O-$ is found. At present there are 9450 such structures in the physical organic databank. That is, compounds with variation on either side of the phenolic $-O-$ are found. Except for the carboxyl group this constitutes the largest class of compounds in the physical organic databank. We are now working on the means for systematically limiting this type of search. Various restrictions can be used in limiting the catch. For example, searching for only compounds with a free phenolic OH yields 2536 hits (there are many OR substituted aromatic rings).

Field 18 contains statistics which, although important for individual equations, are of more interest at present for global evaluation of the QSARs. To obtain some feeling about the quality of the correlations we can use the command **18 r<0.95**, which finds 628 equations in the physical organic database that have r^2 less than 0.9 (18% of the database). Going to the show mode and entering **15** we can peruse the parameters which are involved in these poorer quality equations. It is not surprising to find terms in E_S, B, *P* and **2. Steric effects correlated by E_S and the sterimol parameters B1 and B5 are difficult to correlate. log *P* (found via string search on *P*) has been used to correlate a variety of chromatographic and diffusion processes which also are not generally sharp correlations. The equations with parabolic terms are found using **2. These have been used to deal with nonlinear Hammett equations. Eliminating these and considering only equations based on σ, σ^+, or σ^- we find 3340 examples of which 428 (12.8%) are of lower quality ($r < 0.95$). Setting a higher standard **18 r>0.98** ($r^2 = 0.96$), we find 2148 QSARs or 64% of the database with r^2 greater than 0.96.

It seemed likely to us that the quality of fit might be associated with the size of the set of substituents studied. Often studies by physical organic chemists were limited to small sets of so-called well-behaved groups. This was a consequence of the interest in establishing a value for ρ. Very few laboratories have been interested in extending the range of substituents and its dependence on quality of fit. Considering the above 3340 examples of sets with σ parameters and using two commands: **18 r>0.95** and **18 n<8** isolates all sets based on seven or fewer compounds having $r > 0.95$. We find that 90.6% of the sets with seven or fewer data points have $r > 0.95$. Considering data sets with **n>20** we find that only 75% meet this standard.

Another way to consider quality is by the number of outliers associated with each set. These are marked for each QSAR since it is very important to keep account of them. The search **18 omit>0** finds all sets in which one or more data points have been omitted in deriving the final equation (1340 out of

4000). Setting a lower standard **18 omit>1** finds only 478.

Still another way to compare quality is to compare results from different classes. σ constants are normally defined in water or ethanol−water solvents. Therefore one might expect better results in reactions run in water than say a nonpolar, nonhydrogen bonding solvent such as benzene. The following search isolated high-quality reactions run in water:

	search command	number of sets
1	**1 aqueous**	1702
2	**1 not %**	654
3	**15 not ES P B **2**	329
4	**18 omit <1**	222

Command 2 eliminates mixed solvents. When mixed solvents are used the % of each is indicated. Command 3 removes QSARs with steric terms (ES, B1, and B5 sterimol parameters). The string search on *P* removes QSARs with hydrophobic terms log *P* and π and **2 removes equations with parabolic terms. The **18 omit<1** yields 222 sets where no data points were omitted. The label omit indicates the number (if any) of data points in each set which were not included in deriving the QSAR. In some instances this refers to ortho subtituents for which no effort was made to parameterize the steric effect.

Using the same method for benzene as the solvent we find the following result:

	command	number of sets
1	**1 benzene**	287
2	**1 not %**	279
3	**15 not P ES B **2**	243
4	**18 omit<1**	187

In the case of water out of 329 examples 222 had no data points omitted (67.5%). For benzene out of 243 examples 187 had no data points omitted (77%). The results with benzene are no worse than for water. This is a rough means for comparison and one might want to examine both sets to consider the types of reactions involved (see below).

Quality can also be considered via the correlation coefficient by using the 329 examples where water alone was the solvent. Of these, 292 (87%) have $r > 0.95$ and 220 (75.3%) have $r > 0.98$. In the case of the 243 benzene examples, 214 (88%) have $r > 0.95$ and 168 (69.1%) have $r > 0.98$. Again it is found that results using benzene as the solvent are just as good as those where water was the solvent.

One might expect QSARs based on σ^+ to be more sensitive to an aqueous solvent rather than a nonpolar solvent. Selecting sets from the 329 good examples based on σ^+ with water alone as solvent, 59 are found; of these, 53 have $r > 0.95$ (89.8%) and 42 (71.2%) have $r > 0.98$. Of 75 sets for benzene, 66 have $r > 0.95$ (88%) and 52 (69.3%) have $r > 0.98$. Again, the results with water are no better than with benzene.

However, considering QSARs for reactions run in the vapor phase (72 examples) is a step beyond benzene. Omitting studies of mass spectra (P4N) yields 62 cases. Narrowing this to sets having only σ, σ^+, or σ^- leaves 40 examples. All of these have $r > 0.95$. Placing the further restriction **n > 7** reduces the group to 24 (60%) and **omit > 0** narrows the group to 13 examples (32.5%). From our limited set of examples of vapor-phase reactions, the quality is not as good as with benzene except in terms of *r*.

Still another way of evaluating quality is by the standard deviation. Searching on **18 s <0.2** finds that of the 329 aqueous examples based on σ^+ 263 (80%) have a standard deviation less than 0.2. For the 243 benzene sets 219 (90%) have $s < 0.2$. Of course, none of the above quality tests is absolute, but taken all together they do show that even though σ constants have mostly been defined in water or ethanol−water (σ^+ was defined in 90% acetone/10% water), they correlate reactions in the nonpolar solvent benzene very well.

By means of field 18 we can explore the physical organic database of 4000 sets by the following stringent test:

command	hits	% of database
18 n>10	974	24
18 r>0.98	428	11
18 omit<1	258	6.5
18 s<0.1	126	3.2

A great limitation on quality is that so many studies are restricted to a small number of compounds; the authors' main interest being to establish a value of ρ. Over 500 substituents have had σ_m and σ_p values reported,[15] and it is a pity that more of these have not been studied.

If we lower our standards as follows, a larger catch is obtained:

command	hits	% of database
18 n>7	1870	46.8
18 r>0.95	1488	37.2
18 omit<1	871	21.8
18 s<0.15	615	15.4

To obtain some feeling about the relationship between type of reaction and quality of fit we can ignore the standard deviation limitation and consider the 871 data sets. Requiring a low standard deviation tends to elminate those sets which have a wide range in the dependent variable.

field	hits	total in class	% of class
2 P1	390	1655	23.6
2 P2 P3	159	701	22.7
2 P4	75	261	28.7
2 P5	47	303	15.5
2 P6	120	626	19.2
2 P7	44	284	15.5
2 P8	12	97	12.4
2 P9	27	107	25.2
2 P12	42	242	17.4
2 P13	10	68	14.7

Of the 10 classes, seven (ionization, hydrolysis and solvolysis, spectra, rearrangements, radical reactions, substitution) are about equally well fit with about 20% of their members falling in the well fit class. Four (miscellaneous reactions, addition, elimination, complex formation) are less well fit.

Finally, the quality of Hammett correlations can be analyzed in terms of the deviation of individual data points. The command **19 −0.05<dev<0.05** isolates 558 sets where no member has a calculated value which deviates from the experimental by more than ±0.05. Setting the range at ±0.2 finds 2009 sets. These are *very* stringent tests since the failure of one datapoint to meet the standard eliminates the

set. It is mostly small sets which fall in the high-quality range. For example, of the 558 sets with all members in the range ±0.05, only 57 contain more than eight data points and only 19 have more than 12. This searching mode can be used to find sets with outliers at any level.

Field 11, substituents, can sometimes be of help in studying the behavior of individual groups. Use is somewhat limited since we have not tried to use a standard method of labeling. For example **11 Me** finds 3285 sets in which one or more methyl groups are present. Searching on **11 Me CH3** (the space means we find either Me or CH_3) finds 3772 examples, showing that in 487 instances CH_3 is also used for methyl. In this way we find that 513 sets contain CF_3, nine sets contain SF_5, and two sets contain $OCHF_2$.

From another point of view all QSARs containing a σ^+ term (864) can be sequestered. Now employing **11 OMe OCH3** we find 684 examples, but with **11 NH2** we find only 61 cases. We could now examine these sets to see how OCH_3 and NH_2 are fit. The same technique can be employed with various classes to study the behavior of particular substituents for certain types of reactions.

It has often been suggested that MO calculations based on molecules in a vacuum might, for this reason, be less good for LFER studies than experimentally obtained Hammett constants from reactions in solution. As discussed above the results in the nonpolar benzene compared to highly polar H-bonding water imply that the solvent atmosphere around reactants may not be as important as once suspected. Clearly those studying MO-based parameters can find many ways to compare their results with those obtained via the Hammett equation using the present database-search program. However, it must be remembered that the present study is based on only 4000 QSARs.

The quality of linear free energy relationships has improved significantly in recent years, as one might expect. For the 10 year period 1955−1964, we have 715 QSARs, of which 567 have $r > 0.95$ (79.3%) and 399 have $r > 0.98$ (55.8%). For the period 1984−1993, we have 548 equations, of which 496 have $r > 0.95$ (96.5%) and 374 have $r > 0.98$ (68.2%). We think this same trend will occur with the biological QSAR; however, it is not apparent at present since the better experiment work is offset by more complex chemical structures being studied.

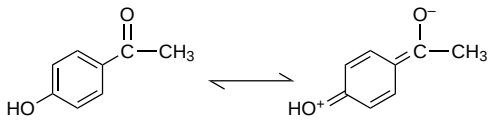
For the biological portion, it is difficult to make comparisons of the statistical types shown above for the physical organic part of the database. In general, the biological equations are much more complex and the variation in the structure of the substituents is much greater. The following search can be compared with that for the physical organic section:

command	hits	% of database
18 n>7	2053	68.4
18 r>.95	799	26.6
18 omit<1	447	14.9
18 s<.15	216	7.2

One of the big differences is that those doing biological work normally test larger sets of chemicals.

Our primary interest is to understand how organic compounds affect living organisms. Chemicals may affect critical receptors in cells or animals by relatively simple reversible interactions which depend on hydrophobic or steric properties (Fischer lock and key fit). Possibly a majority of drugs operate in this manner. However in many instances chemical reactions will be involved. Probably all metabolic processes involve chemical reactions as do many mutagenic and carcinogenic initiators. Strong electrophiles and compounds that become toxic via radical formation also involve bond making and breaking. Of course the confirmed absence of an electronic term in a biological QSAR is of itself important in what it implies for mechanism. Thus it should be possible to increase our understanding of these processes by comparing electronic terms (or their absence) for biochemical studies with those from physical organic chemistry. It is for this reason that we have been constructing the physical organic part of our system. Although 4000 examples of physical organic QSARs is only a fraction of the published examples, we believe that it contains representative examples of most types of organic reactions. We have reported a few such comparisons.[1,11−13]

At this point a brief definition of the Hammett parameters is in order.[15,16] The normal σ for substituents on aromatic systems where strong resonance between substituent and reaction center does not occur is defined as $\sigma = \log K_X - \log K_H$, where K_H is the ionization constant for benzoic acid (normally in water or 50% ethanol) and K_X is that for a substituted benzoic acid. For instances where there is strong resonance interaction between substituent and reaction center two other parameters, σ^- and σ^+, are employed. Of these σ^- is defined using the ionization constants from phenols or anilines similar to σ: $\sigma^- = \log K_X - \log K_H$, where K refers to the ionization of anilines or phenols. This takes into account resonance of the following type:

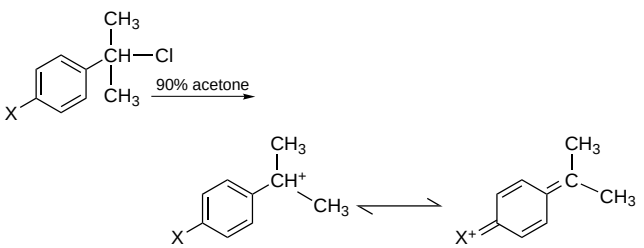


That is, a negative charge is being delocalized. Meta substituents do not exhibit this electronic effect so that we find: $\sigma_m^+ \approx \sigma_m^- \approx \sigma_m$. That most normal σ values show rather little resonance effect is evident from the following QSAR:[15]

$$\sigma_p = 1.19\sigma_m - 0.08$$
$$n = 530,\ r^2 = 0.885,\ s = 0.137 \quad (1)$$

For most (but not all) substituents, σ_m and σ_p are highly collinear.

While σ and σ^- are defined via equlibrium constants, σ^+ is defined by the rate of solvolysis of cumene chlorides in 90% acetone/10% water:



In this instance it is a positive charge which is being delocalized:

$$\sigma^+ = \log (k_X/k_H)/\rho$$

where ρ is found by plotting certain well-behaved meta substituents against σ from the benzoic acid system. In this way a value of 4.54 has been found for ρ which places σ^+ on roughly the same scale as σ. Not only does σ^+ correlate reactions where positive charge is delocalized by substituents, such as aromatic electrophilic substitution, it also often, but not always, correlates free radical reactions. Finally, a fourth σ constant, σ_I or F, is needed to correlate reactions in saturated systems where only the field/inductive effect is involved.[15]

Hammett constants have been astonishingly successful in correlating almost every kind of organic reaction in all sorts of solvents. Eventually one assumes that quantum chemical calculations will replace them, but this is not possible at present. That is, when MO parameters are compared with σ constants the latter generally give better correlations. This is especially true when strong resonance interactions are involved. An interesting aspect of the Hammett equation is that geometry is not taken into consideration. That is, the three models (benzoic acids, phenols, and cumene chlorides) define the geometry for all correlations. In the case of the field/inductive effect geometry is normally not a consideration. MO calculations are usually prefaced with the comment that "all bond lengths and angles are completely optimized". This is the most time-consuming operation in such calculations. Of course the Hammett approach is not able to include as wide a range of structures in a single equation as molecular orbital parameters. The molecular orbital parameters do offer the potential of greater insight since one is able to consider charge densities on all atoms as well as the HOMOs and LUMOs. This plethora of parameters brings up the collinearity problem. Possibly using principal components is the way out, but how this will affect comparative QSAR is not yet evident. Indeed, σ, σ^-, σ^+, and σ_I might be viewed as kinds of principal components. Large numbers of Hammett constants of the various types have been published.[15,17,18]

Although there have been a number of more or less successful attempts to factor σ, σ^-, and σ^+ into resonance and field/inductive components, in our present analysis we shall not attempt to work with the factored parameters. Simpler parameterization makes for easier comparison. Although our system is not well suited at present to study individual data points, it can be done. For example, in the past we listed σ_p^- for NH_2 as -0.15, OH as -0.16, OMe as -0.16, and NMe_2 as -0.12.[18] In order to check this we can isolate all physical organic QSARs with negative ρ by the search **16 −100<" S- "<0**. Examining the 218 sets obtained we find that using $\sigma_p^- = \sigma_p$ for NH_2, OH and OMe is generally better than the more positive values; however, for NMe_2 the value of $\sigma_p^- = -0.83$ does not seem to work as well, and -0.12 *may* be a better choice. It would be valuable to use the present database to obtain a better definition of σ constants, but this would be a large undertaking with our present software.

The solvent plays an important role in organic reactions; accordingly, we have alloted field 1 in the physical organic database for its description. For instance, searching on **aqueous** makes 1701 hits; however, many of these have been done with mixed solvents. The amount of second or third solvent is indicated by a % sign. Now searching on **1 not %** eliminates all examples containing solvents other than water, leaving 654 examples. Searching on **1 ethanol** finds 928 examples. This string search would also find m*ethanol*, methoxy*ethanol*, etc. Searching on **1 " ethanol "** finds only 540. Using the qualification **1 not %** reduces this to 145 examples. However there are a few examples where the % ethanol was not specified. Adding the further qualification **1 not aqueous** reduces the number to 142.

To review the literature on the effect of solvent on a reaction, the ionization of benzoic acid can be used as an example:

	field	command	yield
1	2	**" P1 "**	763
2	12	**SMILES of benzoic acid**	57
3	10	**PKA**	52
4	3	**not misc**	51
5	15	**not ES F**	45

In step 1, quotes are placed on the class P1 for ionization; otherwise, classes containing 1 (10, 15, etc.) would be included. Step 3 eliminates five examples where equilibrium constants (rather than pK_A) are the dependent variable and step 4 eliminates one example of a miscellaneous set of acids in which benzoic acid was included. Step 5 elminates examples containing ortho substituents (correlated by E_S) and a study where σ was factored into F and R. Table 3 lists results of the search in order of increasing value of ρ. Note that if log K had been used instead of pK_A, the sign of ρ would be positive.

Although a number of the sets in Table 3 contained ortho substituents, none of these were included in deriving the equations. While ortho substituents could be included by using additional parameters (steric and field/inductive), we wished to have the simplest possible equations for making comparisons. In sets 8, 10, 14, and 15, 3-hydroxybenzoic was an outlier and was not included. One wonders about the purity of the compound. In set 26 the 3-NO_2 congener is an outlier. In set 41 the COOH substituents were not included in the derivation of the QSAR. In the examples having large values of ρ the quality of the correlation as indicated by r is good; however, the 95% confidence limits on ρ are wide. In examples 42, 44, and 45 where reactions were run under high pressure there seems to be a small, but significant, effect on ρ.

A perusal of the values in Table 3 shows that as the polarity of the solvent increases, the value of ρ approaches that found in water. Even though confidence limits are rather wide on ρ for the nonpolar solvents, where it is possible to compare ρ values the agreement is good. The ρ values from Table 3 yields a rather interesting correlation with the Dimroth−Reichardt solvatochromic parameter E_T^N:[18a,b] ρ =

Table 3

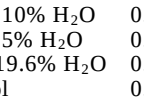
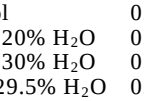
	solvent	E_T^N	T, °C	ρ	n	r
1	acetonitrile	0.46	20	−2.52(±0.36)	9	0.987
2	DMSO	0.44	20	−2.50(±0.43)	7	0.989
3	DMSO	0.44	25	−2.39(±0.36)	11	0.981
4	DMF	0.40	–	−2.39(±0.20)	9	0.996
5	DMSO	0.44	–	−2.37(±0.41)	14	0.964
6	DMSO, 4.5% H_2O	0.48	20	−2.23(±0.38)	9	0.982
7	DMSO, 9% H_2O	0.52	20	−2.11(±0.28)	9	0.989
8	ethanol	0.66	25	−2.00(±0.26)	10	0.988
9	DMSO, 13.6% H_2O	0.55	20	−1.97(±0.27)	9	0.989
10	ethanol, 10% H_2O	0.68	25	−1.92(±0.25)	9	0.989
11	ethanol, 5% H_2O	0.66	25	−1.90(±0.41)	8	0.977
12	DMSO, 19.6% H_2O	0.58	20	−1.87(±0.25)	9	0.989
13	methanol	0.77	25	−1.81(±0.61)	5	0.983
14	ethanol, 20% H_2O	0.70	25	−1.80(±0.21)	9	0.991
15	ethanol, 30% H_2O	0.73	25	−1.73(±0.20)	10	0.980
16	DMSO, 29.5% H_2O	0.63	20	−1.63(±0.20)	9	0.991
17	dioxane, 18% H_2O	0.54*	25	−1.56(±0.15)	17	0.984
18	dioxane, 28.5% H_2O	0.62	25	−1.56(±0.11)	17	0.992
19	tetrahydrofuran, 40% H_2O	0.62	25	−1.56(±0.39)	5	0.991
20	ethanol, 50% H_2O	0.76	25	−1.54(±0.16)	18	0.990
21	dioxane, 45% H_2O	0.69	25	−1.49(±0.10)	17	0.991
22	ethanol, 50% H_2O	0.76	25	−1.48(±0.20)	7	0.993
23	butanol	0.60*	25	−1.47(±0.09)	8	0.998
24	DMSO, 38.8% H_2O	0.69	20	−1.47(±0.17)	9	0.991
25	$HOCH_2CH_2OH$	0.79	25	−1.43(±0.23)	11	0.979
26	ethanol, 60% H_2O	0.80	25	−1.42(±0.29)	7	0.984
27	butyl cellosolve, 50% H_2O		25	−1.40(±0.21)	7	0.991
28	methanol	0.77	25	−1.37(±0.23)	11	0.976
29	DMSO, 48.8% H_2O	0.75	20	−1.34(±0.13)	8	0.995
30	dioxane, 58.5% H_2O	0.77	25	−1.32(±0.10)	17	0.990
31	DMSO, 58.6% H_2O	0.85	20	−1.26(±0.08)	8	0.998
32	DMSO, 68.3% H_2O	0.85	20	−1.16(±0.08)	8	0.998
33	acetone, 75% H_2O	0.82	25	−1.09(±0.09)	16	0.990
34	DMSO, 78.7% H_2O	0.85	20	−1.08(±0.10)	6	0.998
35	H_2O	1.00	35	−1.01(±0.05)	9	0.999
36	H_2O	1.00	15	−1.01(±0.06)	9	0.998
37	H_2O	1.00	25	−1.01(±0.05)	9	0.998
38	H_2O	1.00	20	−1.01(±0.05)	9	0.998
39	H_2O	1.00	30	−1.01(±0.05)	9	0.999
40	H_2O	1.00	45	−1.01(±0.06)	9	0.998
41	methanol, 50% H_2O	0.84	20	−1.01(±0.30)	9	0.949
42	H_2O, 1000 Bars		25	−0.99(±0.04)	7	0.999
43	DMSO, 89.2% H_2O	1.00	20	−0.99(±0.15)	5	0.997
44	H_2O, 2000 Bars		25	−0.96(±0.04)	7	0.999
45	H_2O, 3000 Bars		25	−0.94(±0.04)	7	0.999

2.49(±0.31) E_T^N − 3.40(±0.23), n = 39, r^2 = 0.874, s = 0.172. The ρ and E_T^N values from Table 3 were used to derive this expression. Two of the ρ marked by with an asterisk were poorly fit and not included in the correlation. Although the correlation is not very high, it seems reasonable considering that no correction was made for temperature. This equation could be used to estimate ρ for the ionization of benzoic acids in other solvents since hundreds of E_T^N values have been reported.[18a,b]

From the study of the solvent effect on the ionization of benzoic acids one might expect to find such a solvent effect on reactions with enzymes when ligands are completely engulfed in an active site. So far we have not found any convincing evidence for such an effect.

The above examples give a brief overview of the present system and its potential as a tool for increasing our ability to improve our understanding of structure−activity relationships since the introduction of the Hammett equation in 1935. We hope that our ideas will be useful to others working on the problem of organizing structure−activity studies. We now turn to the specific problem of comparative QSAR of biological data based on σ.

Table 4. Estimated *f* Values for Some Common Atomic Units

group	f	set no. from Table 5[a]
$-CH_2-$	2.19	3,4
$-O-$	1.62	6
$-S-$	1.43	7
$-Se-$	1.16	11
$-CH{=}CH-$	2.16	5
$-C{\equiv}C-$	2.24	18
$-SO-$	2.95	9
$-SO_2-$	1.83	10
	3.27	13
	3.02	15
$-CO-$	1.2	26a

[a] These sets from Table 5 have been used to define f.

Comparison of Biological QSARs by Means of Rho (ρ) Values

For many years, there has been interest among physical organic chemists in comparing ρ values from similar reactions.[19,20] We have been concerned with making such comparisons for biological QSARs.[1,2,11−13] An early generalization of considerable value is that the introduction of a CH_2 moiety between a reaction center and a substituted benzene ring reduces ρ by almost $^1/_2$.[19,20] For example, in the esterification of X-Ar-CH_2COOH and X-Ar-CH_2CH_2COOH by diazodiphenylmethane, ρ is, respectively, 0.40 and 0.22.[20] A general expression for reactions involving a unit change in charge at an atom separated by i units of saturated bonds is $\rho = 2.5/2^i$. Wells[19] listed several examples. We have now extended this relationship in Tables 4 and 5.

In Table 4 we have listed the fragment values f which have been used to calculate the ρ values in Table 5. To estimate $f$$CH_2$, examples 3 and 4 were used since the 95% confidence limits for ρ in example 2 were rather large. Those examples in Table 5 where calculated values are not given have been used to estimate f values. For our purposes, we calculate ρ using the f values in Table 4 as follows:

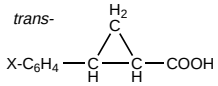
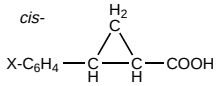
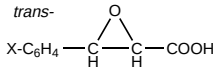
$$\rho = \rho_o/F$$

where ρ_O is from the reference solvent system and F is the product of all f. The third column in Table 4 refers to the set number in Table 5 used to calculate the f value.

The use of the data in Table 4 can be illustrated with example 22 from Table V. To calculate ρ for this example use ρ of −1.54 from example 12 for the ionization of benzoic acid in 50% ethanol. This is then divided by the product of 2.95 (f for SO) and 2.16 (f for −CH=CH−) to yield −0.24.

The values of Table 4 have been used to make estimates of ρ for ionization, esterification and ester hydrolysis for which we have suitable data. However, many other reaction types should be amenable

Table 5. Ionization of Acids (pKa)

	compound	solv	T, °C	ρ observed	ρ calculated	n	r	ref
1	X-C_6H_4COOH	H_2O	25	−1.00				
2	X-$C_6H_4CH_2COOH$	H_2O	25	−0.587(±0.101)	−0.457	5	0.996	22
3	X-$C_6H_4CH_2CH_2COOH$	H_2O	25	−0.203(±0.02)		5	0.999	23
4	X-$C_6H_4CH_2CH_2COOH$	H_2O	25	−0.215(±0.044)		8	0.980	24
5	X-C_6H_4CH=$CHCOOH$	H_2O	25	−0.463(±0.086)		9	0.979	24
6	X-$C_6H_4OCH_2COOH$	H_2O	25	−0.282(±0.044)		26	0.938	25
7	X-$C_6H_4SCH_2COOH$	H_2O	25	−0.319(±0.070)		14	0.944	26
8	X-$C_6H_4SCH_2COOH$	H_2O	20	−0.399(±0.097)	−0.319	12	0.945	27
9	X-$C_6H_4SOCH_2COOH$	H_2O	25	−0.155(±0.047)		13	0.910	26
10	X-$C_6H_4SO_2CH_2COOH$	H_2O	25	−0.249(±0.076)		13	0.908	26
11	X-$C_6H_4SeCH_2COOH$	H_2O	20	−0.394(±0.062)		18	0.959	27
12	X-C_6H_4COOH	50% EtOH	25	−1.54(±0.16)		18	0.990	28,29
13	*trans*- 	50% EtOH	25	−0.471(±0.067)		11	0.982	30
14	*cis*- 	50% EtOH	25	−0.413(±0.011)	−0.471	8	0.966	30
15	*trans*- 	50% EtOH	25	−0.510(±0.085)		5	0.996	31
16	*cis*-X-C_6H_4CH=$CHCOOH$	50% EtOH	25	−0.687(±0.122)	−0.713	8	0.985	32
17	*trans*-X-C_6H_4CH=$CHCOOH$	50% EtOH	25	−0.691(±0.046)	−0.713	7	0.998	32
18	X-C6H4C≡CCOOH	50% EtOH	25	−0.686(±0.064)		10	0.993	33
19	*cis*-X-C_6H_4CH=$CHCH$=$CHCOOH$	50% EtOH	25	−0.337(±0.121)	−0.330	9	0.928	34
20	*trans*-X-C_6H_4CH=$CHCH$=$CHCOOH$	50% EtOH	25	−0.340(±0.129)	−0.330	8	0.935	34
21	*cis*-X-C_6H_4SCH=$CHCOOH$	50% EtOH	25	−0.530(±0.097)	−0.499	6	0.991	35
22	*cis*-X-C_6H_4SOCH=$CHCOOH$	50% EtOH	25	−0.388(±0.087)	−0.242	6	0.987	35
23	*cis*-X-$C_6H_4SO_2CH$=$CHCOOH$	50% EtOH	25	−0.319(±0.122)	−0.390	6	0.964	35
24	*trans*-X-C_6H_4SCH=$CHCOOH$	50% EtOH	25	−0.650(±0.138)	−0.499	6	0.988	35
25	*trans*-X-C_6H_4SOCH=$CHCOOH$	50% EtOH	25	−0.282(±0.067)	−0.242	6	0.986	35
26	*trans*-X-$C_6H_4SO_2CH$=$CHCOOH$	50% EtOH	25	−0.330(±0.085)	−0.390	6	0.983	35
26a	X-C_6H_4C(=O)CH_2CH_2COOH	50% EtOH	25	−0.226(±0.04)		13	0.979	35a
		Esterification						
27	X-C_6H_4COOH (with diazodiphenylmethane)	EtOH	30	0.941(±0.191)		4	0.998	36
28	X-$C_6H_4CH_2COOH$ (with diazodiphenylmethane)	EtOH	30	0.435(±0.194)	0.430	4	0.989	36
29	X-$C_6H_4CH_2CH_2COOH$ (with diazodiphenylmethane)	EtOH	30	0.227(±0.022)	0.196	4	0.999	36
30	X-$C_6H_4OCH_2COOH$ (with diazodiphenylmethane)	EtOH	30	0.240(±0.108)	0.265	4	0.989	36
		Hydrolysis						
31	X-$C_6H_4COOC_2H_5$	87.83% EtOH	30	2.506(±0.157)		18	0.993	37−39
32	X-$C_6H_4CH_2COOC_2H_5$	87.83% EtOH	30	0.824(±0.267)	1.144	6	0.974	38
33	X-$C_6H_4CH_2CH_2COOC_2H_5$	87.83% EtOH	30	0.477(±0.392)	0.522	4	0.965	38,39
34	X-C_6H_4CH=$CHCOOC_2H_5$	87.83% EtOH	30	1.329(±0.086)	1.160	14	0.995	38
35	X-C_6H_4C≡$CCOOC_2H_5$	87% EtOH	10	1.072(±0.078)	1.119	6	0.999	40
36	X-$C_6H_4COOC_2H_5$	85% EtOH	25	2.548(±0.202)		12	0.989	41
37	X-$C_6H_4CH_2CH_2COOC_2H_5$	78% EtOH	30	0.528(±0.077)	0.531	20	0.959	42
38	X-C_6H_4CH=$CHCOOC_2H_5$	85% EtOH	25	1.295(±0.094)	1.180	13	0.994	43
39	*trans*-X-C_6H_4CH=$CHCOOC_2H_5$	88% EtOH	30	1.281(±0.101)	1.180	8	0.997	44
40	*cis*-X-C_6H_4CH=$CHCOOC_2H_5$	88% EtOH	30	1.054(±0.088)	1.180	6	0.998	44

to this treatment so that the method can be used to uncover unusual correlations such as those for the aniline mustards shown in sets 1−3 of Table 6.

In Table 4, two types of fragments have been defined. Fragments such as CH_2 and cyclopropyl do not contain lone pair electrons and hence transmit electronic effects by the field/inductive mode. For others, especially −CH=CH− and −C≡C−, transmission via resonance will be important. Despite this, mixing of fragments (such as in examples 21−26) gives reasonable results. This may be due to the fact that all of our examples are based on the ionization of carboxylic acids. As eq 1 shows, most substituents do not interact strongly via resonance with the carboxyl group.

The problem when strong through resonance is applied can be illustrated with the following set of germanium derivatives and their ρ^+ values where the $GeEt_3$ group is cleaved by acid. After attack by H^+, the positive charge is strongly delocalized by substituent X, which correlates with σ^+.

	ρ^+	f	**calcd** ρ^+ (f = 1.67)
1 X-$C_6H_4GeEt_3$	−4.03(±0.52)		
2 XC_6H_4C≡$CGeEt_3$	−2.44(±0.60)	1.65	−2.41
3 XC_6H_4(C≡C)$_2GeEt_3$	−1.59(±0.20)	1.53	−1.44
4 XC_6H_4(C≡C)$_3GeEt_3$	−0.86(±0.14)	1.84	−0.86
		1.67 mean value	

The confidence limits on ρ^+ are rather wide; nevertheless, there is fair agreement for the three fragment values. They are significantly smaller than the 2.24 in Table 4 for −C≡C−. Unfortunately, we have nothing to compare with this interesting study by Eaborn et al.[44a]

The value of Tables 4 and 5 can be illustrated with the examples of QSAR eqs 2−9 below for cytotoxic compounds which have been extensively studied as antitumor agents.[13] Equations 2−9 show how these substances react with nucleophiles, models for their reaction with DNA. It is presumed that their anticancer action is due to their ability to couple two strands of DNA. Except for eqs 2 and 9, there is a rather narrow range in ρ (−1.84 to −2.38; mean = −2.04). Why eq 2 is so much different is not clear, but the selenium analog in eq 9 is expected to be different. It is a bit surprising that the sulfur analog in eqs 7 and 8 has ρ values so near to the nitrogen analogs. It should be noted that in a few instances using σ^- gives slightly better results. For the purposes of comparison we have used σ for all examples. Whether or not ρ is influenced by σ or σ^- is determined by two factors. If the choice of substituents is poor (i.e. few substituents for which σ values differ significantly from σ^-) the correlation may be essentially the same. If the electronic demands on the substituent by the reaction center in the transition state are not strong, then there will be little difference between ρ and ρ^-.

nucleophilic substitution of
X-$C_6H_4N(CH_2CH_2Cl)_2$ in 50/50 acetone/water
66 °C with nitrobenzylpyridine[13]

$$\log k = -3.21(\pm 0.84)\sigma - 2.91 \qquad n = 7,\ r^2 = 0.95 \quad (2)$$

substitution of X-$C_6H_4N(CH_2CH_2Cl)_2$
in ethanol 80 °C with nitrobenzylpyridine[13]

$$\log k = -2.38(\pm 0.99)\sigma - 2.24 \qquad n = 9,\ r^2 = 0.82 \quad (3)$$

hydrolysis of X-$C_6H_4N(CH_2CH_2Cl)_2$ in 50/50
acetone/water 66 °C[13]

$$\log k = -2.31(\pm 1.1)\sigma - 3.40 \qquad n = 6,\ r^2 = 0.879 \quad (4)$$

half-life in fetal calf serum
X-$C_6H_4N(CH_2CH_2Cl)_2$[13]

$$\log T_{1/2} = -1.84(\pm 0.45)\sigma - 0.48 \qquad n = 11,\ r^2 = 0.922 \quad (5)$$

hydrolysis of X-$C_6H_4N(CH_2CH_2Cl)_2$ in 50/50
acetone water 66 °C[13]

$$\log k = -1.84(\pm 0.40)\sigma - 4.10 \qquad n = 11,\ r^2 = 0.923 \quad (6)$$

nucleophilic substitution of X-$C_6H_4SCH_2CH_2Br$
in 4 to 1 acetone buffer 30 °C with
nitrobenzylpyridine[13]

$$\log k = -1.97(\pm 0.16)\sigma - 4.10 \qquad n = 12,\ r^2 = 0.987 \quad (7)$$

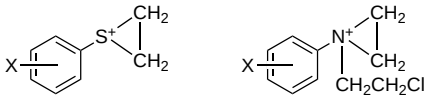
hydrolysis of X-$C_6H_4SCH_2CH_2Br$ in 5 to 4
acetone buffer 40 °C[13]

$$\log k = -1.87(\pm 0.37)\sigma - 0.55 \qquad n = 9,\ r^2 = 0.952 \text{ (omitted: 3-OMe)} \quad (8)$$

nucleophilic substitution of X-$C_6H_4SeCH_2CH_2Cl$
in 50% acetone 37 °C with nitrobenzylpyridine[13]

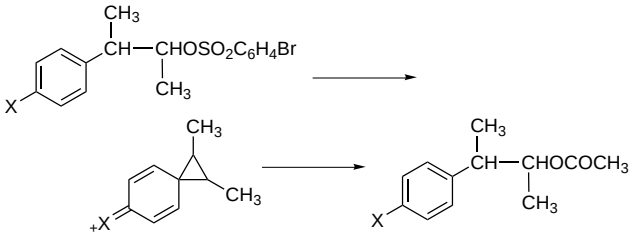
$$\log k = -1.29(\pm 0.48)\sigma - 2.08 \qquad n = 5,\ r^2 = 0.961 \quad (9)$$

In the light of the falloff in ρ as CH_2 moieties are placed between substituents (Table 5) and the reaction center, the high values for ρ in eqs 2−9 would seem surprising. In the case of the aniline mustards and their sulfur and selenium analogs, three atoms stand between the ring and the halogens undergoing displacement. The explanation for the large ρ's is that these reactions depend on the availability of the lone pair electrons on N, S, or Se to displace the halogen to form onium intermediates:



These intermediates then rapidly react with the nucleophiles (water, nitrobenzylpyridine, or those in the biosystem).

Another such example is the following solvolysis in acetic acid-acetic anhydride which is best correlated by σ^+:[45]



Lone pair electrons on the substituent promote the displacement of $BrC_6H_4SO_2O^-$ to form the intermediate which then reacts with acetic anhydride to yield the acetate. For this reaction, ρ^+ is −2.46(±0.36), which is not far from that found for the above reactions. In the examples of the antitumor-type compounds an S_N2 type reaction appears to occur with the lone pair electrons on the heteroatom displacing the leaving group. In the case of the butyl brosylates an S_N1 type reaction seems to occur with the carbocation attacking an electron-rich site on the aromatic ring.

Another interesting example similar to the above is the solvolysis of X-$C_6H_4CH_2CH_2CH_2CH_2Hg^+ClO_4^-$ where the correlation is with σ^+ with ρ of −0.90(±0.14).[45a] If the action were through four CH_2 bonds, ρ would have to be almost 20 for $C_6H_4Hg^+$-ClO_4, according to Table 5. It seems likely that a cyclic intermediate is involved.

It is important not to ignore the confidence limits on ρ. Generally, as ρ increases, so do the confidence limits (we have used 95% confidence limits).

Table 6. Biological QSAR Containing a Term in σ

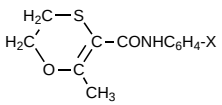
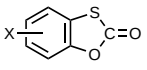
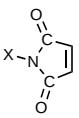
set	system	compound	ref
1	I_{50}, Chinese hamster UV4 cells $\log 1/C = -2.54(\pm0.25)\sigma + 5.78$ omitted: 4-NO_2	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 13$, $r^2 = 0.978$	48
2	I_{50}, Chinese hamster UV4 tumor cells $\log 1/C = -2.23(\pm0.49)\sigma + 5.80$	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 11$, $r^2 = 0.920$	49
3	I_{50}, Chinese hamster cells AA8 $\log 1/C = -2.16(\pm0.77)\sigma + 5.02$	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 6$, $r^2 = 0.938$	50
4	acylation by acetyltransferase $\log V_{max}/K_{m=} -2.10(\pm0.64)\sigma + 0.66(\pm0.34)\log P - 1.81(\pm0.65)I - 2.12$ omitted: 2-Me $I = 1$ for ortho substituents	X-C_6H_4-NH_2 $n = 16$, $r^2 = 0.843$	51
5	25% increase in life span of mice with B-16 melanoma $\log 1/C = -2.07(\pm0.48)\sigma + 0.37(\pm0.28)C\log P - 0.92(\pm0.44)\log(\beta 10^{ClogP} + 1) + 3.36$ omitted: 4-CN; 4-$CH_2CH(NH_2)COOH$; 4-SO_2NH_2	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 20$, $r^2 = 0.889$	52
6	I_{50}, P388 leukemia cells $\log 1/C = -2.00(\pm1.3)\sigma + 5.66$	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 6$, $r^2 = 0.824$	50
7	180% increase in life span of mice with P-388 leukemia $\log 1/C = -1.98(\pm0.83)\sigma - 0.31(\pm0.08)C\log P + 0.53(\pm0.44)I + 5.13$ omitted: 2-$CH(NH_2)CH_2COOH$; 2-OCH_2CH_2COOH; 4-$CH_2C(NH_2)(Me)COOH$ $I = 1$ for 2-substituents	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 15$, $r^2 = 0.907$	53
8	Inh., *V. maydis* $\log 1/C = -1.93(\pm0.76)\sigma + 5.85$ omitted: 2-Cl	 $n = 8$, $r^2 = 0.864$	54
9	Inh., rat liver alcohol dehydrogenase $\log 1/K_{i=} -1.80(\pm0.77)\sigma + 1.22(\pm.16)\log P + 4.87$	4-X-pyrazoles $n = 14$, $r^2 = 0.970$	55
10	Inh., ethanol oxidation in rat hepatocytes $\log 1/K_{i=} -1.80(\pm0.87)\sigma + 1.27(\pm.33)\log P - 0.20(\pm0.12)(\log P)^2 + 4.75$	4-X-pyrazoles $n = 14$, $r^2 = 0.941$	55
11	Inh., horse alcohol dehydrogenase $\log 1/K_i = -1.73(\pm0.99)\sigma + 1.64(\pm0.39)\pi - 0.41(\pm0.29)B5 + 1.30$ omitted: 4-NMe_2	4-X-pyrazoles $n = 15$, $r^2 = 0.903$	56
12	Inh., *T. mentagrophytes* $\log 1/C = -1.75(\pm0.67)\sigma + 3.74(\pm1.80)\log P - 0.54(\pm0.30)(\log P)^2 + 2.19$ omitted: 2-naphthoxy	 X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 9$, $r^2 = 0.927$ optimum $\log P = 3.4$	57
13	increase in lifespan of mice with B-16 melanoma $\log 1/C = -1.70(\pm0.53)\sigma$ - $0.42(\pm0.13)\log P + 5.52$ omitted: 3,5-di-NHCOMe; 3,5-di-NH_2	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 20$, $r^2 = 0.845$	58
14	LD_{50}, rats $\log 1/C = -1.68(\pm0.42)\sigma + 0.68(\pm0.30)I - 0.67(\pm0.35)I_2 + 3.73$ omitted: NHCOMe, Br; $(CH_2)_3COOH$, Cl $I = 1$ for Br, 0 for Cl or I; $I_2 = 1$ for COOH	4-X-$C_6H_4N(CH_2CH_2Y)_2$ $n = 16$, $r^2 = 0.886$	53
15	deacylation of acylchymotrypsin $\log K_2 = -1.60(\pm0.58)\sigma + 0.88$	X-C_6H_4-NH_2 $n = 8$, $r^2 = 0.881$	59
15a	Inh., endocytosis in erythrocyte ghosts $\log 1/C = -1.60(\pm0.61)\sigma + 0.66(\pm0.21)C\log P + 3.10$	 $n = 9$, $r^2 = 0.925$	60
16	ED_{90}, rat with Walker 256 tumor $\log 1/C = -1.59(\pm0.50)\sigma + 0.88(\pm0.36)I - 0.53(\pm0.42)I_2 + 3.81$ omitted: NHCOMe, Cl; $(CH_2)_3COOH$, Cl $I = 1$ for Y = Br, 0 for Y = Cl or I; $I_2 = 1$ for COOH	4-X-$C_6H_4N(CH_2CH_2Y)_2$ $n = 16$, $r^2 = 0.848$	53
17	25% increase in life span of mice with L1210 leukemia $\log 1/C = -1.51(\pm0.73)\sigma - 0.34(\pm0.08)C\log P + 0.70(\pm0.30)I + 5.12$ omitted: 4-COOH; 4-OCH_2COOEt $I = 1$ for ortho substituents	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 19$, $r^2 = 0.896$	53
18	Inh., *p*-hydroxybenzoate hydrolase $\log 1/C = -1.47(\pm0.43)\sigma + 4.30$	X-C_6H_4COOH $n = 6$, $r^2 = 0.956$	61

Table 6 (Continued)

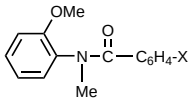
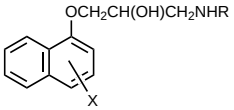
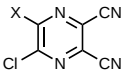
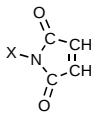
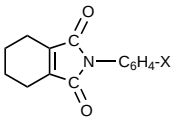
set	system	compound	ref
19	Inh., appressorial penetration of fungus in cellophane film		62
	$\log 1/C = -1.35(\pm0.48)\sigma - 0.66(\pm0.17)E_S\text{-}2 + 0.62(\pm0.30)C\log P - 2.88 \log(\beta 10^{\text{Clog}P} + 1) + 1.98$	$n = 21$, $r^2 = 0.955$; optimum $\log P = 4.3(\pm0.35)$	
20	I_{50}, Red spider (Titelarius) eggs $\log 1/C = -1.36(\pm0.61)\sigma + 1.37\ (\pm0.38)I + 4.44$ omitted: 3-COOMe; 3-Me, 4-OMe $I = 1$ for COOR	X-$C_6H_4SCH_2CH_2Br$ $n = 14$, $r^2 = 0.852$	63
21	25% increase in lifespan of mice with L1210 leukemia $\log 1/C = -1.30(\pm0.48)\sigma - 0.28(\pm0.11)C\log P + 4.44$	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 11$, $r^2 = 0.920$	52
22	local anesthesia, guinea pig $\log 1/C = -1.25(\pm0.53)\sigma + 0.65(\pm0.36)\log P + 2.03$	X-$C_6H_4COOCH_2CH_2NEt_2$ $n = 8$, $r^2 = 0.895$	64
23	25% increase in life span of mice with P-388 leukemia $\log 1/C = -1.20(\pm0.41)\sigma - 0.36(\pm0.08) + 5.76$ omitted: 3-CH_2CH_2COOH; 3-$CH(NH_2)CH_2COOH$; 4-$(CH_2)_3CHO$	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 19$, $r^2 = 0.869$	53
24	40% reduction, maximum driving frequency in rabbit atria		65
	$\log 1/C = -1.08(\pm0.69)\sigma + 0.56(\pm0.18)\log P' + 4.44$ omitted: R = $(CH_2)_3COOH$, X = H log P' is the distribution coefficient at pH 7.4	$n = 11$, $r^2 = 0.877$	
25	Inh., trypsin; human $\log 1/K_i = -1.07(\pm0.62)\sigma + 1.29(\pm1.3)\text{B1-2} - 0.75(\pm0.39)\text{B5-2} + 4.59$ omitted: 3-COOH B1 and B5 are sterimol parameters for ortho substituents	X-$C_6H_4C(=NH)NH_2$ $n = 13$, $r^2 = 0.812$	66
26	papain hydrolysis	X-$C_6H_4NHCOCH_2CH$-$(NHCOMe)CH_2C_6H_5$	67
	$\log 1/K_i = -1.04(\pm0.24)\sigma + 1.11$	$n = 6$, $r^2 = 0.974$	
27	I_{50}, barnyard grass		68
	$\log 1/C = -0.89(\pm0.20)\sigma + 0.93(\pm.21)\pi - 1.46(\pm0.37)\log(\beta 10^{\pi} + 1) - 0.54$	$n = 21$, $r^2 = 0.894$	
28	rabbit anesthesia $\log\text{RBR} = -0.85(\pm0.29)\sigma + 0.26(\pm0.21)C\log P - 1.10$ omitted: $CH_2CH_2CH_3$; NHC_2H_5	X-$C_6H_4COOCH_2CH_2NEt_2$ $n = 11$, $r^2 = 0.857$	69
29	Inh., trypsin $\log 1/K_i = -0.83(\pm0.22)\sigma + 0.22(\pm0.08)\pi' - 3 + 0.60(\pm0.18)I\text{-M} + 0.22(\pm0.04)\text{MR4} + 0.59(\pm0.23)I\text{-}1 + 0.36(\pm0.21)I\text{-}2 + 0.52(\pm0.17)I\text{-}3 + 0.99$ I-M = 1 for meta substituents; I-1 = 1 for $-CH{=}CHCOOR$; I-2 = 1 for $-CH_2COOR$; I-3 = 1 for $-CH_2CH_2COR$ or $-OCH_2COR$	X-$C_6H_4C(=NH)NH_2$ $n = 110$, $r^2 = 0.805$	70
30	I_{50}, ATP-ase		60
	$\log 1/C = -0.71(\pm0.66)\sigma + 0.51(\pm0.20)\pi + 2.50$	$n = 10$, $r^2 = 0.845$	
31	dissociation of cytochrome P-450 $\log K_2/K_1 = -0.68(\pm0.37)\sigma - 1.03(\pm0.24)\pi + 0.99$	X-$C_6H_4C(Me)_2COOH$ complex $n = 6$, $r^2 = 0.984$	71
32	mutagenesis of *S. typhimurium* TA1535 $\log 1/C = -0.62(\pm0.35)\sigma + 0.28(\pm0.08)\pi + 1.28$ omitted: H; 4-$CONHC_3H_7$	X-$C_6H_4N(CH_2CH_2Cl)_2$ $n = 13$, $r^2 = 0.875$	72
33	I_{50}, root of sawa millet		73
	$\log 1/C = -0.60(\pm0.35)\sigma + 1.77(\pm0.32)L - 0.31(\pm0.08)L^2 - 0.95(\pm0.21)B4\text{-}4 + 4.07$	$n = 28$, $r^2 = 0.865$	
34	mouse liver conjugation of glycine $\log k = -0.58(\pm0.18)\sigma + 0.54(\pm0.13)\log P - 0.52(\pm0.17)\text{MR-}3,4 + 5.40$ omitted: 4-NMe_2; 3-Cl, 4-NO_2	X-C_6H_4COOH $n = 19$, $r^2 = 0.874$	74

Table 6 (Continued)

set	system	compound	ref
35	I_{50}, chloroplasts, spinach photosystem II	X-$C_6H_4NHC(=O)R$, X-$C_6H_4NHCONH_2$, X-$C_6H_4NHCOOR$	75
	$\log 1/C = -0.57(\pm0.34)\sigma + 1.08(\pm0.16)\log P - 0.50(\pm0.13)L(Y)_R + 0.29(\pm0.18)W_1(Y)_R - 0.95(\pm0.19)T_R(Y)_R - 2.21(\pm0.39)T_R(Y)\text{-PH} - 0.32(\pm0.14)W_R(X_1) + 1.02(\pm0.23)I_U + 8.16$ omitted: 5 data points	$n = 69$, $r^2 = 0.849$	
36	I_{50}, chloroplasts, soybean photosystem II $\log 1/C = -0.56(\pm0.48)\sigma + 0.97(\pm0.24)\log P + 3.17$ omitted: 3-Et; 3-NO_2; 4-F	X-$C_6H_4NHCON(Me)OMe$ $n = 19$, $r^2 = 0.838$	76
37	dissociation of cytochrome P-450 $\log K_2/K_1 = -0.53(\pm0.24)\sigma - 0.83(\pm0.16)\pi + 1.17$ omitted: 4-Br	X-$C_6H_4CH_2OOH$ complex $n = 11$, $r^2 = 0.951$	71
38	I_{50}, *E. coli* $\log 1/C = -0.48(\pm0.24)\sigma + 0.61(\pm0.13)C\log P - 1.56(\pm0.40)\log(\beta 10^{ClogP} + 1) + 4.45$	X-$C_6H_4SO_2C_6H_4$-4-NH_2 n= 17, $r^2 = 0.912$	77
39	D-amino acid oxidase oxidation $\log k = -0.46(\pm0.24)\sigma + 0.30(\pm0.14)C\log P + 1.70$ omitted: NMe_2	X-$C_6H_4CH_2CH(NH_2)COOH$ $n = 7$, $r^2 = 0.914$	78
40	Inh., *M. tuberculosis* $\log 1/C = -0.45(\pm0.27)\sigma + 0.11(\pm0.12)\log P + 3.71$ omitted: 4-OH	4-X-$C_6H_4CONHNH_2$ $n = 8$, $r^2 = 0.835$	79
41	I_{50}, chloroplast spinach photosystem II $\log 1/C = -0.44(\pm0.49)\sigma + 0.98(\pm0.18)\log P - 0.89(\pm0.48)\log(\beta 10^{\log P} + 1) - 0.14(\pm0.06)\text{Br} + 3.24$ Br = factor for branching of 4-substituents	X-$C_6H_4NHCONH(Me)OCH_3$ $n = 37$, $r^2 = 0.885$	80
42	mouse anticonvulsants $\log 1/C = -0.43(\pm0.20)\sigma - 0.29(\pm0.18)C\log P + 0.52(\pm0.27)\text{B1-2} + 0.34(\pm0.13)\text{B5-6} + 2.95$ omitted: 2-Cl-4,6-di-Me	X-$C_6H_4NHCONH$-4-pyridyl $n = 19$, $r^2 = 0.848$	81
43	dissociation of cytochrome P-450 $\log K_2/K_1 = -0.26(\pm0.14)\sigma - 0.91(\pm0.08)\pi + 0.50$ omitted: 4-OEt	X-C_6H_4COOOH complex $n = 16$, $r^2 = 0.980$	71
44	I_{50}, chloroplasts, spinach photophosphorylation $\log 1/C = 0.35(\pm0.13)\sigma + 0.45(\pm0.06)\pi + 3.87$	X-$C_6H_4NHCH_2CH_2$-NEt_2 $n = 19$, $r^2 = 0.945$	82
45	subtilisin hydrolysis $\log 1/K_m = 0.41(\pm0.08)\sigma + 0.20(\pm0.07)\pi\text{-3} + 0.24(\pm0.08)\text{MR-4} + 3.80$	X-$C_6H_4OCOCH_2NHCOC_6H_5$ $n = 30$, $r^2 = 0.850$	83
46	papain hydrolysis $\log k_{cat}/K_m = 0.43(\pm0.17)\sigma + 0.38(\pm0.07)\pi\text{-4} + 4.60$ omitted: 4-C_6H_5; 4-NHCOMe; 4-OC_4H_9	X-$C_6H_4C(=O)NCH_2$-COO-3-pyridyl $n = 22$, $r^2 = 0.857$	84
47	chloroplasts, spinach; Inh., of photophosphorylation $\log 1/C = 0.43(\pm0.07)\sigma + 0.28(\pm0.14)C\log P + 3.11$	X-$C_6H_4NHC_6H_4$-X′ $n = 15$, $r^2 = 0.951$	85
48	β-acetylglucosaminidase hydrolysis $\log 1/K_m = 0.44(\pm0.14)\sigma - 0.34$ omitted: 2,4-di-NO_2	β-X-phenylacetylaminoglucose $n = 10$, $r^2 = 0.876$	86
49	growth inh., *M. audovinii* $\log 1/C = 0.45(\pm0.38)\sigma + 0.41(\pm0.17)C\log P + 2.93$ omitted: 2-COOH	X-$C_6H_4NHCH_2C_6H_3$-2-OH, 5-Cl $n = 8$, $r^2 = 0.980$	87
50	Inh., *S. aureus* $\log 1/C = 0.48(\pm0.19)\sigma + 0.69(\pm0.10)\log P - 1.07(\pm0.85)\log(\beta 10^{\log P} + 1) + 0.86$ omitted: 2-Me; 6-Cl; 4-NH_2; 2,4-di-NH_2; 4-OH; 2-$CH_2CH{=}CH_2$ optimum $\log P = 4.5(\pm0.9)$	phenols $n = 33$, $r^2 = 0.916$	88
51	antagonist, β-receptor in guinea pig atria $\log 1/C = 0.47(\pm0.32)\sigma + 0.66(\pm0.14)C\log P - 4.29(\pm1.1)\log(\beta 10^{ClogP} + 1) + 4.62$ omitted: 4-$NHSO_2Me$, 3-C_6H_5, 4-C_6H_5, 3,4-$(CH)_4$	X-$C_6H_4CH(OH)CH_2NHCHMe_2$ $n = 22$, $r^2 = 0.856$	89
52	carboxypeptidase A hydrolysis $\log 1/K_m = 0.53(\pm0.27)\sigma + 0.44(\pm0.27)\text{MR} + 1.86$	4-X-$C_6H_4COOCH(Et)COOH$ $n = 6$, $r^2 = 0.931$	90
53	papain hydrolysis $\log 1/K_m = 0.55(\pm0.20)\sigma + 0.61(\pm0.08)\,\pi'_3 + 0.45(\pm0.11)\text{MR-4} + 2.00$	X-$C_6H_4OC(=O)CH_2NHSO_2Me$ $n = 32$, $r^2 = 0.893$	91
54	papain hydrolysis $\log 1/K_m = 0.56\,(\pm0.19)\sigma + 1.08(\pm0.23)\,\pi'_3 + 0.67(\pm0.27)\text{MR-4} + 3.75$ omitted: 3-SO_2NH_2; 4-SO_2NH_2; 3,5-di-OMe	X-$C_6H_4OC(=O)CH_2NHCOC_6H_5$ $n = 24$, $r^2 = 0.861$	92
55	papain hydrolysis $\log 1/K_m = 0.56(\pm0.19)\sigma + 0.57(\pm0.26)\text{MR} + 1.92(\pm0.15)I + 1.82$	X-$C_6H_4OCOCH_2NH$-R R = SO_2Me or COC_6H_5 $n = 20$, $r^2 = 0.981$ $I = 1$ for R = COC_6H_5	93

Table 6 (Continued)

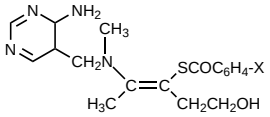
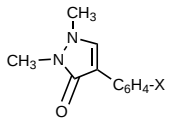
set	system	compound	ref
56	human erythrocytes, absorption		94
	logRBR = 0.56(±0.12)σ + 0.18	$n = 8$, $r^2 = 0.955$	
57	I_{50}, chymotrypsin log1/(I/S) = 0.56(±0.29)σ + 0.82(±0.21)*C*log *P* - 3.16	X-$C_6H_4OCH_2COCH_3$ $n = 15$, $r^2 = 0.870$	95
58	ficin hydrolysis log 1/K_m = 0.57(±0.16)σ + 0.84(±0.14)π'-3 + 0.41(±0.18)MR-4,5 + 3.60	X-$C_6H_4OCOCH_2NHCOC_6H_5$ $n = 28$, $r^2 = 0.885$	96
59	LD_{30}, *E. coli* log 1/*C* = 0.61(±0.38)σ + 0.75(±0.21)*C*log *P* + 0.91	X-$C_6H_4CH{=}CH_2$ $n = 7$, $r^2 = 0.966$	97
60	I_{50}, ATP-transferase rat liver log 1/*C* = 0.61(±0.13)σ + 0.62(±0.11)π + 1.62	X-$C_6H_4OCH_2CH_2CH(NH_2)COOH$ n =11, $r^2 = 0.980$	98
61	ficin hydrolysis log 1/K_m = 0.62(±0.21)σ + 0.55(±0.11)π'-3 + 0.28(±0.12)MR-4 + 2.67 omitted: 3-C_6H_5	X-$C_6H_4OCOCH_2NHSO_2CH_3$ $n = 32$, $r^2 = 0.817$	99
62	bromelain D hydrolysis log k_o/K_m = 0.63(±0.09)σ + 0.46(±0.12)MR + 2.21	4-X-$C_6H_4OCOCH_2NHCOC_6H_5$ $n = 8$, $r^2 = 0.992$	100
63	I_{50}, *A. niger* log 1/*C* = 0.63(±0.39)σ + 1.04(±0.20)log *P* − 1.06(±0.45)log(β10$^{\log P}$ + 1) + 0.76	phenols $n = 18$, $r^2 = 0.956$ optimum log *P* = 4.2(±4.2)	101
64	bromelain B hydrolysis log k_o/K_m = 0.68(±0.31)σ + 0.60(±0.33)MR + 1.16	4-X-$C_6H_4OCOCH_2NHSO_2Me$ $n = 8$, $r^2 = 0.955$	100
65	bleaching activity on radish log 1/*C* = 0.70(±0.41)σ + 0.59(±0.16)*C*log *P* − 0.38(±0.18)E_S-3 + 2.40 omitted: 3-OC_6H_5	X-$C_6H_4NHCON(Me)OMe$ $n = 17$, $r^2 = 0.931$	102
66	bromelain B hydrolysis log k_o/K_m = 0.70(±0.37)σ + 0.50(±0.41)MR + 2.62	4-X-$C_6H_4OCOCH_2NHCOC_6H_5$ $n = 8$, $r^2 = 0.925$	100
67	Inh., *S. aureus* log 1/*C* = 0.70(±0.26)σ + 0.34(±0.12)*C*log *P* + 4.15 omitted: 2-COOH, 2-NH_2	X-C_6H_4SH $n = 9$, $r^2 = 0.955$	103
68	I_{50}, tetrahymena log 1/*C* = 0.72(±0.29)σ + 0.51(±0.17)log *P* + 1.97	methyl- and chloroanilines $n = 30$, $r^2 = 0.924$	104
69	I_{50}, cholinesterase, rat log 1/*C* = 0.73(±0.56)σ − 0.40(±0.21)E_S-3 + 3.41	X-$C_6H_4CH{=}CHC_5H_4N^+CH_3$ $n = 10$, $r^2 = 0.884$	105
70	actinidin hydrolysis log 1/K_m = 0.74(±0.15)σ + 0.50(±0.13)π'-3 + 0.25(±0.21)MR-4 + 2.90 omitted: 4-SO_2NH_2, 3,5-di-OMe	X-$C_6H_4OCOCH_2NHCOC_6H_5$ $n = 27$, $r^2 = 0.860$	106
71	LD_{50}, rice stem borer log 1/*C* = 0.75(±0.45)σ − 0.40(±0.20)B5 + 1.69(±0.44)π-sum − 0.18(±0.13)(π-sum)^2 − 1.17(±0.30)I + 5.68 omitted: X = 4-N=NC_6H_5, Y = Cl; X = 4-OC_2H_5, Y = Cl I = 0 for 2,6-di-F and 1 for 2,6-di-Cl	4-X-C_6H_4-NHCONHCO-C_6H_3-2,6-di-F (2,6-di-Cl) $n = 29$, $r^2 = 0.832$	107
72	rat, matriuretic action log 1/*C* = 0.75(±0.21)σ − 0.18(±0.16)*C*log *P* + 0.32 omitted: 3-NO_2, 4-Cl; 3-NO_2, 4-CF_3	X-$C_6H_4SO_2NH_2$ $n = 14$, $r^2 = 0.859$	108
73	Inh., carbonic anhydrase log 1/K_i = 0.77(±0.21)σ + 0.30(±0.21)π + 0.71(±0.30)E_S-2 + 0.31 omitted: 2-NO_2	X-$C_6H_4SO_2NH_2$ $n = 18$, $r^2 = 0.910$	109
74	chymotrypsin hydrolysis log K_2/K_m = 0.78(±0.35)σ + 0.76(±0.45)MR-3 + 0.71	4-$OHC_6H_4CH_2CH$-(NHCOCH$_3$)CO-NHC_6H_4-X $n = 7$, $r^2 = 0.922$	110
75	10% elongation of avena coleoptile log 1/*C* = 0.78(±0.63)σ + 1.25(±0.29)log *P* + 3.50 omitted: 3-Pr	X-$C_6H_4CH_2COOH$ $n = 15$, $r^2 = 0.910$	111
76	I_{50}, photosystem II in pea chloroplasts log 1/*C* = 0.81(±0.39)σ + 0.71(±0.23)π + 4.47	 $n = 11$, $r^2 = 0.947$	112

Table 6 (Continued)

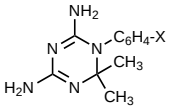
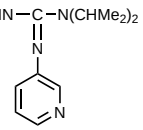
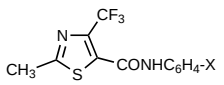
set	system	compound	ref
77	LD_{50}, chinese hamster cells $\log 1/C = 0.89(\pm0.16)\sigma + 0.59(\pm0.14)\log P' + 1.26$ P' indicates partitioning was done in buffer	phenols $n = 9$, $r^2 = 0.984$	113
78	Inh., influx of phenylalanine in rabbit intestine in vitro $\log 1/K_i = 0.90(\pm0.37)\sigma + 0.18(\pm0.20)\pi + 1.80$	X-$C_6H_4CH_2CH(NH_2)$-COOH $n = 8$, $r^2 = 0.935$	114
79	I_{50}, L5178y tumor cells $\log 1/C = 0.90(\pm0.56)\sigma + 1.41(\pm0.23)\pi' - 1.65(\pm0.26)\log(\beta 10^{\pi'} + 1) + 0.51(\pm0.20)I - 0.25(\pm0.24)\mathrm{OR} + 0.63(\pm0.32)\mathrm{DO} + 7.95$ π' indicates $\pi = 0$ for 3-OR and 4-OR, DO = 1 for alkyl groups, OR = 1 for OR groups omitted: 3-CH(Me)C_6H_5	 $n = 65$, $r^2 = 0.819$	115
80	Inh., carbonic anhydrase $\log 1/C = 0.90(\pm0.23)\sigma + 0.23(\pm0.17)C\log P + 5.37$ omitted: 2-CH_3; 2-Cl; 2-NO_2	X-$C_6H_4SO_2NH_2$ $n = 16$, $r^2 = 0.927$	116
81	Inh., of sulfate exchange in erythrocyte membrane $\log 1/C = 0.90(\pm0.42)\sigma + 0.34(\pm0.25)\pi + 2.05$	X-C_6H_4-SO_3^- $n = 11$, $r^2 = 0.880$	117
82	Inh., mouse tumor dihydrofolate reductase $\log 1/K_i = 0.90(\pm0.68)\sigma + 1.19(\pm0.21)\pi' - 1.38(\pm0.27)\log(\beta 10^{\pi'} + 1) + 0.50(\pm0.24)I + 6.20$ omitted: 3-CN; 3-COOEt	(see compounds of set **80**) $n = 38$, $r^2 = 0.875$	118
83	arylsulfatase hydrolysis $\log 1/K_m = 0.93(\pm0.30)\sigma + 2.52$	X-C_6H_4-OSO_3H $n = 10$, $r^2 = 0.867$	119
84	Inh., carbonic anhydrase $\log \mathrm{RBR} = 0.94(\pm0.53)\sigma + 1.23(\pm0.29)\pi + 2.51$ omitted: 2-Me	X-$C_6H_4SO_2NH_2$ $n = 12$, $r^2 = 0.922$	120
85	oat coleoptiles, 10% increase in elongation $\log 1/C = 0.97(\pm0.51)\sigma + 0.95(\pm0.32)L - 5.54(\pm1.3)\log(\beta 10^{L} + 1) + 1.26(\pm0.20)\pi + 1.39$ omitted: 3,4-$(CH)_4$, $NHCOCH_3$, C_4H_9, $NHCOC_6H_5$, C_6H_5, OCF_3	3-X-C_6H_4-OCH_2COOH $n = 19$, $r^2 = 0.951$	121
86	Inh., carbonic anhydrase $\log 1/K_i = 0.98(\pm0.33)\sigma + 0.53(\pm0.10)C\log P - 0.33(\pm0.11)\mathrm{B\text{-}5,3} + 6.10$	X-$C_6H_4SO_2NH_2$ $n = 32$, $r^2 = 0.851$	122
87	Inh., fungus, *Sphaerotheca fulginea* $\log 1/C = 0.98(\pm0.17)\sigma + 0.46(\pm0.15)\mathrm{HB} + 3.13$ HB = 1 for hydrogen bonding substituents	 $n = 13$, $r^2 = 0.943$	123
88	Inh., dihydrofolate reductase $\log 1/K_i = 1.05(\pm0.74)\sigma + 0.56(\pm0.11)\pi' - 0.59(\pm0.15)\log(\beta 10^{\pi'} + 1) + 0.61(\pm0.22)I + 5.73$ omitted: 3-CH_3; 3-$CO_2C_2H_5$	compounds of set **80** $n = 62$, $r^2 = 0.808$	124
89	β-D-xylosidase hydrolysis $\log V = 1.01(\pm0.17)\sigma + 0.19(\pm0.18)\mathrm{MR} - 0.03$ omitted: 4-F, 3-NO_2	4-X-C_6H_4-β-D-xylo-pyranosides $n = 15$, $r^2 = 0.932$	125
90	I_{50}, adenosine deaminase $\log(1/S)0.5 = 1.09(\pm0.43)\sigma + 0.30(\pm0.17)C\log P - 0.35$ omitted: COMe	4-(3-X-benzyl)adenines $n = 9$, $r^2 = 0.927$	126
91	rat liver esterase hydrolysis $\log K\text{-rel} = 1.11(\pm0.20)\sigma - 0.15$ omitted: 3-Me	X-$C_6H_4COOC_6H_4$-4-NEt_2 $n = 10$, $r^2 = 0.953$	127
92	I_{50}, mitochondrial succinate dehydrogenase $\log 1/C = 1.11(\pm0.27)\sigma + 0.80(\pm0.16)C\log P + 1.27(\pm.35)\mathrm{I} + 3.43$ $I = 1$ for 2,6-disubstituted omitted: 2-Me, 4-Br, 6-Me; 2,4,6-tri-Br, 3-OH; 2,4-di-Br, 6-COMe; 2,4-di-Br, 6-COOEt; 2,4-di-Cl, 6-Br	 $n = 42$, $r^2 = 0.887$	128

Table 6 (Continued)

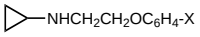
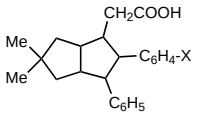
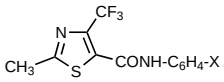
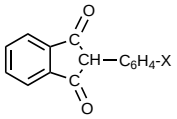
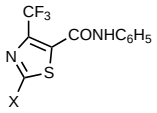
set	system	compound	ref
93	rat liver microsomal oxidation to CO_2 $\log k = 1.13(\pm0.33)\sigma + 1.58$ omitted: 4-OH; 4-NO_2	X-methylene-dioxybenzenes $n = 9$, $r^2 = 0.903$	129
94	proteinase K hydrolysis $\log k_{cat}/K_m = 1.09(\pm0.49)\sigma + 0.37(\pm0.47)\mathrm{MR} + 1.60$ omitted: COOH	4-X-C_6H_4OCOMe $n = 8$, $r^2 = 0.916$	130
95	I_{50}, tetrahymena $\log 1/C = 1.24(\pm0.36)\sigma + 0.80(\pm0.12)\log P + 1.45$	phenols $n = 27$, $r^2 = 0.897$	131
96	I_{50}, monoamine oxidase, rat liver		132
	$\log 1/C = 1.28(\pm0.49)\sigma + 0.45(\pm0.31)\pi + 0.79(\pm0.21)E_S\text{-}3,5 + 3.81$	$n = 18$, $r^2 = 0.901$	
97	β-D-glucosidase, almond meal hydrolysis $\log k_{cat} = 1.30(\pm0.33)\sigma - 0.53(\pm0.21)\pi + 1.81$	X-C_6H_4-glucosides $n = 13$, $r^2 = 0.941$	133
98	Inh., *E. coli* $\log 1/C = 1.31(\pm0.23)\sigma - 0.24(\pm0.14)C\log P + 5.28$ omitted: 4-NMe_2	4-H_2N-$C_6H_4SO_2NHC_6H_4$-4′-X $n = 16$, $r^2 = 0.927$	134
99	MIC, *S. mutans* $\log 1/C = 1.44(\pm0.70)\sigma + 0.83(\pm0.22)\log P' - 1.10(\pm0.27)\mathrm{MR} + 3.81$	5-X-8-OH-quinolines $n = 13$, $r^2 = 0.924$ P' corrected for ionization	135
100	cholinesterase, horse serum hydrolysis $\log V_{max} = 1.56(\pm0.48)\sigma + 0.50(\pm0.32)E_S\text{-}2 - 13.5$ omitted: 2-F, 4-Me	X-$C_6H_4COOCH_2CH_2N^+Me_3$ $n = 13$, $r^2 = 0.879$	136
101	I_{50}, cyclooxygenase, bovine		137
	$\log 1/C = 1.53(\pm0.95)\sigma - 2.40(\pm1.0)\mathrm{B1,2} - 0.36(\pm0.34)C\log P + 10.7$	$n = 13$, $r^2 = 0.812$	
102	I_{50}, succinate dehydrogenase, mitochondria		128
	$\log 1/C = 1.54(\pm0.56)\sigma + 0.78(\pm0.16)\pi + 4.22$ omitted: penta-Cl	$n = 16$, $r^2 = 0.895$	
103	human carbonic anhydrase binding $\log K = 1.55(\pm0.36)\sigma + 0.65(\pm0.08)\log P - 2.13(\pm0.21)I - 3.28(\pm0.21)I\text{-}2 + 6.92$ $I = 1$ for *m*-COOR; I-2 = 1 for *o*-COOR	X-$C_6H_4SO_2NH_2$ $n = 29$, $r^2 = 0.984$	138
104	Inh., seminal bovine vesicles synthesis of prostaglandin		139
	$\log 1/C = 1.56(\pm0.38)\sigma + 0.37(\pm0.09)C\log P + 2.46$ omitted: 4-C_6H_5	$n = 23$, $r^2 = 0.850$	
105	I_{50}, glutamic acid dehydrogenase $\log S/I = 1.62(\pm1.0)\sigma + 0.44(\pm0.28)C\log P - 1.96$ omitted: 5-NHCOMe	salicylic acids $n = 10$, $r^2 = 0.856$	140
106	deacylation at pH 2.07 $\log k = 1.78(\pm0.58)\sigma - 2.06$	X-benzoylchymotrypsin $n = 7$, $r^2 = 0.927$	141
107	deacylation at pH 8.5 $\log k = 1.73(\pm0.38)\sigma - 3.48$ omitted: 4-F; 4-NO_2	X-benzoylchymotrypsin $n = 11$, $r^2 = 0.922$	142
108	I_{50}, mitochondria, rat liver oxygen consumption $\log 1/C = 1.75(\pm0.69)\sigma + 0.82(\pm0.27)\pi + 4.95$	2-X,4-NO_2,6-*sec*-butyl phenols $n = 9$, $r^2 = 0.914$	143
109	I_{50}, thiopurine methyltransferase $\log 1/C = 2.22(\pm1.1)\sigma + 1.30(\pm0.58)\pi\text{-}3 + 3.98$	X-C_6H_4COOH $n = 11$, $r^2 = 0.890$	144
110	carbonic anhydrase, bovine, hydrolysis $\log k = 2.20(\pm0.46)\sigma + 2.62$ omitted: 4-C_6H_5; 4-COO^-	X-C_6H_4OCOMe $n = 7$, $r^2 = 0.968$	145
111	Inh., *E. coli* $\log 1/C = 2.27(\pm0.63)\sigma + 4.31$ omitted: 4-NO_2	X-$C_6H_4N{=}C{=}S$ $n = 8$, $r^2 = 0.927$	146

Table 6 (Continued)

set	system	compound	ref
112	I_{50}, succinate dehydrogenase, mitochondria		128
	$\log 1/C = 2.23(\pm1.1)\sigma + 1.24(\pm1.03)B1 + 4.44$ omitted: CF_3	$n = 8$, $r^2 = 0.942$	
113	deacylation $\log k_3 = 2.77(\pm0.63)\sigma - 0.62$	X-benzoylpapain $n = 8$, $r^2 = 0.951$	147

In summary, unless one has some understanding of the information conveyed by the type of ρ (its sign and magnitude), one works under a handicap in the design of new and more effective congeners as well as in understanding reaction mechanisms. It is for these reasons that we have worked to build an easily searchable database of QSAR based on the traditional parameters. Such QSARs are of value even when using ultra modern 3-D QSAR (e.g. CoMFA) for the insight they provide about mechanism. Indeed, we have found classical QSAR to be highly useful in deriving CoMFA models.

Discussion of Biological QSARs in Tables 6 and 7

We now turn to QSARs in Tables 6 and 7. In the former we have selected all biological QSARs from our database that contain a term in σ (not σ^-, σ^+, σ inductive or F or σ^2) which meet the quality standards of $n > 5$ and $r > 0.895$ ($r^2 > 0.80$). In addition, we have not considered examples where ρ or ρ^- is between -0.30 and 0.33. These values are so small that they provide little mechanistic information. In a number of instances a set of compounds has been tested on several different organisms with essentially the same result. When these are from the same laboratory we have listed only one example. This procedure produces the two diverse sets of QSARs in Tables 6 and 7 (σ^- sets >113), which we believe to be of considerable help in our efforts to wring more mechanistic insight out of the reactions of organic compounds with life and its many components. We plan to do a similar study on QSARs based on σ^+.

At present, we must note that explanations for sign and magnitude of ρ in all examples cannot be offered. What is true for ρ is also true for the coefficients of other parameters in QSAR. Because of collinearity problems, QSARs are often published with unreasonable parameters. Simply having an idea of what to expect can guide researchers to better results.

For example, if we select all biological QSARs having a term in either π or log P and eliminate all those having terms parabolic or bilinear we find 1593 examples. Of these, only 91 have coefficients (h) greater than 1.2 and only 51 (3%) have h greater than 1.4. Of a total of 190 examples where h is negative only 12 (6%) have slopes below -1.2. Hence when one derives a QSAR with h outside the range of ±1.2 some special consideration is called for. The term may be modeling something more than a conventional hydrophobic effect.

Another rough generalization about log P or π is the principle of minimal hydrophobicity.[46] Because of their toxic effects and ease of P450 oxidation, drugs should be made as hydrophilic as possible commensurate with efficacy.

In Tables 6 and 7 the QSARs have been ordered according to increasing size of ρ. In Table 6, ρ ranges from -2.54 to 2.77 and in Table 7 the range is -1.26 to 2.94. Hence finding biological QSAR outside these ranges becomes of special interest. We now consider similar QSARs via set numbers.

Over the years, since the initial work by Everett, Roberts, and Ross,[47] and still today, investigations have been made on the so-called aniline mustards [X-$C_6H_4N(CH_2CH_2Y)_2$] for their antitumor activity. (Note that set numbers are given in bold type; equation numbers, in normal type.) Sets **1**, **2**, **3**, **6** with ρ of -2.54, -2.23, -2.16, and -2.00 (mean = -2.2) are based on such studies. These values of ρ correlate rather well with those for the reactions with nucleophiles (eqs 3−8), mean = -2.03. This is a rather simple situation in which cell toxicity is the end point. In the examples where the goal is to extend the life of an animal [sets **5** (-2.07), **7** (-1.98), **13** (-1.70), **16** (-1.59), **17** (-1.51), **21** (-1.30), **23** (-1.20)] the mean value of ρ (-1.62) is considerably lower. The work with animals is much more difficult because there is a narrow range between the lethal dose and the curative dose of these toxic substances. Comparison of sets **14** and **16** for work from the same laboratory shows that the QSARs for LD_{50} and ED_{90} are almost identical! As eqs 2−9 show, this class of compound reacts readily with nucleophiles, even those as weak as water. Although there is evidence that the mustards link strands of DNA, this cannot be the only way tumor cells are inhibited. Strong nucleophiles such as SH and NH_2 would also be attacked throughout the body.

What is abundantly clear is that there is a parallel between the *in vitro* reaction with nucleophiles (eqs 2−8) and the *in vivo* toxicity. In addition there seems to be some selectivity for tumor cells compared to normal cells. There is more than one way in which the mustards can react with DNA as set **32** shows. In producing mutations there is a low dependence on electron release by substituents ($\rho = -0.62$). Again this raises the question: is it only the reaction with DNA that is important for antitumor activity or are other nucleophiles involved? Another interesting aspect of the aniline mustard-type compound is that their QSAR show low or negative dependence on hydrophobicity. This sometimes is true of other electrophiles.[13]

Set **20** shows that the toxicity of X-$C_6H_4SCH_2CH_2$-Br is similar to that of the mustards. The *in vivo* ρ of -1.36 is smaller than that for the reaction with a

Table 7. Biological QSAR Containing a Term in σ^-

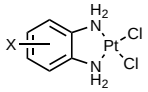
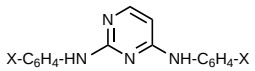
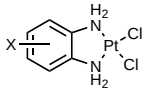
set	system	compound	ref
114	25% increase in life span of mice with B-16 melanoma		148
	$\log 1/C = -1.26(\pm0.72)\sigma^- - 0.80(\pm0.62)\pi + 5.41$ omitted: 4-OMe	$n = 8$, $r^2 = 0.837$	
115	I_{50}, acetyl transfer by choline acetylase $\log 1/C = -1.16(\pm0.72)\sigma^- + 1.16(\pm0.56)\pi + 4.76$ omitted: 2-Me	X-$C_6H_4CH{=}CH$-4-$C_5H_4N^+Me$ $n = 9$, $r^2 = 0.875$	149
116	acetylcholinesterase, eel, hydrolysis, pH 7, 26 °C $\log k_{cat} = -0.81(\pm0.30)\sigma^- + 4.07$	X-C_6H_4OCOMe $n = 8$, $r^2 = 0.880$	150
117	chymotrypsin hydrolysis $\log 1/K_m = 0.41(\pm0.08)\sigma^- + 0.28(\pm0.05)\pi'\text{-}3 + 3.86$	X-$C_6H_4OCOCH_2NHCOC_6H_5$ $n = 28$, $r^2 = 0.897$	151
118	inhibition of alkaline phosphatase $\log 1/K_i = 0.49(\pm0.19)\sigma^- + 2.41$ omitted: H	X-$C_6H_4OPO_3H$ $n = 6$, $r^2 = 0.925$	152
119	I_{50}, *S. fecalis*		153
	$\log 1/C = 0.57(\pm0.34)\sigma^- + 0.30(\pm0.12)\log P + 4.48$	$n = 8$, $r^2 = 0.930$	
120	complex formation with phenol sulfotransferase $\log 1/K_m = 0.70(\pm0.44)\sigma^- + 0.92(\pm0.18)\log P - 1.48(\pm0.38)$MR-4 $- 0.63(\pm0.44)$MR-3 $+ 1.04(\pm0.63)$MR-2 $+ 4.03$ omitted: 4-C_6H_5	phenols $n = 35$, $r^2 = 0.901$	154
121	trypsin hydrolysis at pH 7 $\log k_{cat}/K_m = 0.85(\pm0.12)\sigma^- + 2.15$	X-$C_6H_4OCOCH_2NHCOC_6H_5$ $n = 8$, $r^2 = 0.982$	155
122	I_{50}, folate synthesis, *M. lufu* $\log 1/C = 0.86(\pm0.16)\sigma^- + 4.84$	4-H_2N-$C_6H_4SO_2NHC_6H_4$-4′-X $n = 11$, $r^2 = 0.940$	156
123	I_{50}, mitochondrial oxidative phosphorylation $\log 1/C = 0.87(\pm0.17)\sigma^- + 1.01(\pm0.17)C\log P + 0.51$ omitted: penta-F	phenols $n = 22$, $r^2 = 0.929$	157
124	ED_{50}, mice with malaria, *P. vinckei* $\log 1/C = 1.00(\pm0.63)\sigma^- + 0.72(\pm0.67)C\log P - 1.19$	X-$C_6H_4NHCONH$-C(=NH)NH_2 $n = 7$, $r^2 = 0.841$	158
125	hydrolase, alcolase hydrolysis pH 7.0 $\log k_o/K_m = 0.98(\pm0.14)\sigma^- + 1.83$ omitted: 3-NO_2	X-C_6H_4OCOMe $n = 10$, $r^2 = 0.969$	159
126	Inh., fly head acetylcholinesterase hydrolysis $\log k_2 = 1.35(\pm0.29)\sigma^- + 1.45$	4-X-$C_6H_4SP(=O)(NH_2)OC_2H_5$ $n = 8$, $r^2 = 0.956$	160
127	Inh., *S. cerevisiae* $\log 1/C = 1.43(\pm0.81)\sigma^- + 0.56(\pm0.33)\log P + 2.76$ omitted: 3-NO_2-5-Cl; 3,5,6-tri-Cl	salicylaldehydes $n = 12$, $r^2 = 0.925$	161
128	LD_{50}, flies $\log 1/C = 1.84(\pm0.31)\sigma^- - 0.35(\pm0.18)C\log P + 2.08(\pm0.46)I + 6.89$ $I = 1$ for meta substituents which are either Me or Et	X-$C_6H_4OP(=O)(Me)OEt$ $n = 28$, $r^2 = 0.916$	162
129	LD_{50}, flies $\log 1/C = 1.97(\pm0.47)\sigma^- + 2.94$ omitted: 4-Cl	X-$C_6H_4OP(=O)(Et)OMe$ $n = 6$, $r^2 = 0.970$	163
130	LD_{50}, flies $\log 1/C = 2.13(\pm0.53)\sigma^- + 0.69(\pm0.76)$MR-3 $+ 4.35$ omitted: 4-Et	X-$C_6H_4OP(=S)(C_6H_5)$O-Et $n = 11$, $r^2 = 0.927$	164
131	mitochondria uncoupling oxidative phosphorylation $\log 1/C = 2.18(\pm0.22)\sigma^- + 1.01(\pm0.24)C\log P + 0.27$ omitted: 4-COMe; 2,6-di-I, 4-NO_2	phenols $n = 22$, $r^2 = 0.966$	165
132	mutation of *S. typhimurim* (Ames test)		166
	$\log 1/C = 2.23(\pm0.32)\sigma^- + 5.78$	$n = 13$, $r^2 = 0.956$	
133	LD_{50}, female flies $\log 1/C = 2.28(\pm0.41)\sigma^- - 0.16(\pm0.15)E_S\text{-}3 + 2.71$ omitted: 4-COOH	X-$C_6H_4OP(=O)(OEt)_2$ $n = 8$, $r^2 = 0.976$	167
134	I_{50}, fly head cholinesterase $\log 1/C = 2.45(\pm0.54)\sigma^- - 0.56(\pm0.20)E_S\text{-}3 + 4.82$	X-$C_6H_4OP(=O)(OEt)_2$ $n = 13$, $r^2 = 0.926$	168

Table 7 (Continued)

set	system	compound	ref
135	I_{50}, fly head cholinesterase log $1/C = 2.42(\pm0.49)\sigma^- + 4.31$	4-X-$C_6H_4OP(=O)(Et)OMe$ $n = 7$, $r^2 = .970$	163
136	LD_{50}, thrips log $1/C = 2.56(\pm0.85)\sigma^- + 1.24$ omitted: 3-CMe_3; 4-COOH; 4-CHO	X-$C_6H_4OP(=O)(OEt)_2$ $n = 8$, $r^2 = 0.901$	167
137	Inh., fly head cholinesterase log $K_e = 2.61(\pm0.52)\sigma^- + 0.77(\pm0.35)$B5,4 + 0.84 omitted: 4-Cl; 4-NMe_2	4-X-3,5-xylenyl-di-Et-phosphates $n = 10$, $r^2 = 0.962$	169
138	I_{50}, fly cholinesterase log $1/C = 2.94(\pm0.74)\sigma^- - 0.61(\pm0.55)E_S$-3 + 4.78	4-X-C_6H4OP(=O)(Me)O-Et $n = 15$, $r^2 = 0.893$	162

nucleophilic reagent (eq 7). The substituent effect *in vivo* is not as important as with the isolatednucleophile. This might suggest a more reactive *in vivo* nucleophile (the more reactive the nucleophile the less assistance is required by the substituents).

In the case of the selenium analog (eq 9), ρ is out of line with the other equations (eqs 2−8). Still, it is considerably higher than one would expect from Table 5. For the *in vivo* reaction of these compounds a QSAR with a σ term could not be formulated. A very weak correlation with π was found.

Turning now to set **4** for the acylation of anilines by acetyltransferase we find a large ρ (−2.1) for the overall reation.[51] This is a bit smaller than for the ionization of anilines ($\rho = -2.8$). Enzymic reactions involve at a minimum two steps: complex formation ($1/K_m$) and a catalytic step (k_{cat}). Biochemists have been loath to consider these independently since K_m and k_{cat} are not truly independent variables. We have found, however, that from an SAR point of view more insight can be gained from considering three QSARs. For instance, we have found examples where a substituent effect in the $1/K_m$ step is cancelled in the k_{cat} step. In the case of the P450 demethylation of X-$C_6H_4N(CH_3)_2$, ρ^- of 0.63 in the $1/K_m$ step is cancelled by ρ^- of −0.68 in the k_{cat} step so the overall reaction k_{cat}/K_m is correlated by log P alone.[1] From another point of view, we found that in the enzymatic hydrolysis of phenylglucosides the positive hydrophobic term in the $1/K_m$ step was cancelled by a negative hydrophobic term in the k_{cat} step. Often, as in the present case, substituent effects in the k_{cat} steps are small.

$$\log 1/K_m = -1.55(\pm0.54)\sigma + 0.57(\pm0.28)\log P - 1.39(\pm0.55)I - 2.22 \quad (10)$$

$n = 15$, $r^2 = 0.832$ omitted: 2-CH_3, 3-CH_3

$$\log V_{max} = -0.58(\pm0.11)(\sigma^-)^2 - 0.18(\pm0.13)I + 0.40 \quad (11)$$

$n = 15$, $r^2 = 0.832$ omitted: 2-CH_3, 4-CN

In set **4**, ρ is in rough agreement with the sum of the values in eqs 10 and 11, showing that it is the V_{max} step that is not of overriding importance. Perusal of the physical organic database shows that the transferase reaction has a lower ρ than the pK_a of anilines. The correlation of an extensive set of anilines (51) with their pK_a values in water yields ρ^- of −2.8(±0.08), $r^2 = 0.99$ (ρ^- and ρ^+ refer to correlations by σ^- and σ^+). Other such studies yield similar values for ρ. An interesting QSAR for comparison is that for complex formation of anilines with Ag^+ in 59% ethanol−41% water, where $\rho = -2.1$ (±.58), $n = 9$, $r^2 = 0.91$.[170] As in set **4**, correlation is best with σ. Possibly the most important step in the acylation of anilines is complex formation with an electron-deficient center in the enzyme. However, we have come to expect lower ρ values for enzymic reactions than for the corresponding reactions of organic compounds (see below).

Set **8** is of interest because of its large negative ρ (−1.93). We have no precedent (Table 5) for expecting this could imply that the critical electronic effect of X is on the heterocyclic ring. In fact, there seems to be nothing essential about this ring since it can be replaced with a variety of others without loss of activity [e.g. $CH_3C_6H_4$− or 2,4-$(CH_3)_2$-C-C_3NS−(2,4-dimethyl-1,3-thiazol-5-yl)]. ρ is similar to that found for the aniline mustards. The most likely point for increasing the electron density would appear to be the NH unit. Could this be involved in reaction with an oxidizing system, or could the increase in electron density simply inhibit loss of pesticide effect by hydrolysis? There is some evidence from the work of White and Thorn[171] that these substances inhibit succinic dehydrogenase. From their studies on the I_{50} of 2,6-dichlorophenolindophenol with V. Maydis mitochondria, we have formulated the following QSAR for congeners like those of set **8**:

$$\log 1/C = -0.48(\pm0.77)\sigma - 2.32(\pm1.1)\text{B1,4} - 1.39(\pm0.83)I + 8.4 \quad (12)$$

$n = 15$, $r^2 = 0.781$ omitted: 4-Cl; 2-Me, 4-Cl
I = 1 for 2,6-substitution

In eq 12 the σ term is quite marginal and the B1,4 term (B1,4 is the B1 sterimol parameter for 4-substituents) is by far the dominant term, even though it accounts for only two data points. Hence, one wonders if succinic dehydrogenase is really involved in the fungicidal activity of set **8**. If succinic dehydrogenase is indeed involved in the action of the carboxin fungicides, it must be a form other than that present in the mitochondria. Sets **102** and **112** for similar compounds inhibiting succinic dehydrogenase are quite different. QSAR can cast doubt on a proposed mechanism as well as lend support.

Sets **9**−**11** for the inhibition of alcohol dehydrogenase by 4-X-pyrazoles are interesting in that the values of ρ obtained for pure enzyme, −1.80 and −1.73, are close to the value obtained from inhibition

of the enzyme in liver cells (−1.80). Set **10**, in the case of the liver cells there is a parabolic dependence on hydrophobicity, which is not seen with the isolated enzyme. It is known that these inhibitors react with a positively charged zinc atom in the enzyme. Hence, it makes sense that electron release by substituents increases inhibitory potency.[6] At present, we have no QSAR for pyrazoles from physical organic chemistry to compare with these enzymatic reactions. However, there are other QSARs showing the importance of electron releasing substituents.[172]

inhibition alcohol dehydrogenase, horse liver by $X\text{-}CH_2CONH_2$

$$\log 1/K_i = -0.83(\pm 0.21)\sigma^* + 0.98(\pm 0.39)\log P + 3.69 \quad (13)$$

$$n = 14,\ r^2 = 0.878 \qquad \text{omitted: di-}C_2H_5\text{, tri-}CH_3$$

In eq 13, σ^* is the Taft field/inductive parameter, hence its ρ^* cannot be directly compared with that from σ; however, the negative sign shows that increasing the electron density on the carbonyl moiety increases its affinity for zinc (inhibition constant $1/K_i$).

The QSAR of set **12** has too few data points for the number of terms. Nevertheless, it does provide an estimate of ρ (−1.75), which provides a clue for the mechanism of action of this interesting class of pesticides. Virtually nothing has been published on the chemistry of these hetereocycles.

Set **15** for the deacylation of chymotrypsin by anilines has a ρ (−1.60) similar to that for the acylation of anilines (set **4**) by acetyltransferase (−2.10), as one might expect. Although these values are not similar by the standards of physical organic chemistry they are reasonable for biological QSAR in that they are not far from the mean value of −1.85. The acetyltransferase QSAR has two more terms, indicating a more complex receptor interaction. The acyl group on chymotrypsin does not appear to contact the enzyme in the deacylation step (see also sets **106** and **107**). The only term in these equations is an electronic term which suggests no hydrophobic or steric contact with the enzyme.

Set **15a** correlating the inhibition of endocytosis is interesting in that it has been suggested that this is due to the inhibition of ATPase. In support of this, the authors derived the QSAR in set **30**. Although the hydrophobic effect is essentially the same in QSARs for sets **15a** and **30**, as indicated by the similar coefficients for the hydrophobic terms (π and C log P), the ρ's are different. This is probably the result of a poor selection of substituents in terms of σ. The authors suggest that it is a reaction with an SH group which is important in inhibition of the ATPase. In terms of ρ, the QSAR of set **30** is similar to that of set **33**; however, there is no hydrophobic term in the QSAR of set **33**. Hence, these compounds may be operating by a different mechanism. Because of the negative ρ, the authors suggest that inhibition may involve hydrogen bonding to a carbonyl moiety, which would be promoted by electron-releasing substituents. There is other evidence that the maleimides react with SH groups to inhibit various enzymic systems. For each of the following QSARs, the authors present evidence that the inhibition is due to reaction with sulfhydryl groups.

inhibition of papain hydrolysis by maleimides[173]

$$\log k = 0.53(\pm 0.12)C\log P + 1.56 \qquad n = 8,\ r^2 = 0.952 \quad (14)$$

inhibition of lactate dehydrogenase by maleimides[174]

$$\log k_2 = 0.43(\pm 0.09)C\log P + 0.52 \qquad n = 9,\ r^2 = 0.948 \quad (15)$$

inhibition of D amino acid oxidase by maleimides[175]

$$\log k = 0.37(\pm 0.02)C\log P + 1.16 \qquad n = 6,\ r^2 = 0.999 \quad (16)$$

In each of these examples, only alkyl groups were present on nitrogen. Since there is almost no variation in σ for these substituents, no electronic terms appear in the QSAR. The dependence on log P is similar to that in sets **15a** and **30**. The QSAR of set **33** is quite different because of a lack of a log P term. In this QSAR, L is the sterimol length parameter. One might suspect that L and log P are collinear and hence L might mask a hydrophobic effect, but L and log P are not collinear.

Set **18** for the inhibition of hydroxybenzoate hydrolase by benzoic acids has a ρ of −1.47(±0.43). This is not far from $\rho = -1.0$ for the ionization (pK_a) of benzoic acids in water. It would suggest that the less ionized the acid the more effective it is as an inhibitor. Thus, it would seem that the negatively charged carboxylate does not favor binding.

Equation 17, although it is not a very good correlation (too few data points), is similar to set **18** in that it combines a negative ρ for benzoic acids inhibiting an enzyme.

inhibition of histidine decarboxylase by $X\text{-}C_6H_4COOH$[176]

$$\log 1/C = -0.92(\pm 0.54)\sigma + 0.59(\pm 0.44)\text{B1,4} + 0.40(\pm 0.57)\text{MR} + 1.06 \quad (17)$$

$$n = 11,\ r^2 = 0.823 \qquad \text{omitted: 4-I, 3-OH}$$

These two QSARs suggest that ionization does not favor inhibition. Hydrophobicity is not significant, but the steric parameters imply a complex interaction. Set **34**, which also contains a negative ρ for the benzoic acids, is expected to be more complicated because the study was performed in mouse liver. The log P term probably accounts for transport to the active site. The less ionized benzoic acids would penetrate the liver cells more readily.

Set **19** correlates growth inhibition of fungus by 4-substituted benzanilides. At present, we have little with which to compare this QSAR. Worth mentioning, however, is complex formation of simple benza-

mides (X-$C_6H_4CONH_2$) with BF_3 in tetrahydrofuran, which yields eq 18.[177]

$$pK = -1.62(\pm 0.08)\sigma + 1.72 \qquad n = 12,\ r^2 = 0.994 \tag{18}$$

The substituent effect on polarization of the carbonyl group parallels that of set **19** (K is the dissociation constant; pK is a measure of the basicity of the carbonyl group in complex formation). However, here too, as in the case of set **8**, increase in electron density on N might inhibit hydrolysis. The ρ is similar to that of the benzanilides and might indicate that polarization of the carbonyl bond is of importance in fungicidal activity.

Set **22** is quite similar to the benzanilides (set **19**) and again one wonders about polarization of the carbonyl moiety. It has been suspected that a negatively charged oxygen on the carbonyl group might aid in the binding of these local anesthetics to a receptor. It is also possible that increasing the electron density on the carbonyl group would prevent hydrolysis and thus increase potency, since these esters are readily hydrolyzed *in vivo*. Either or both mechanisms may be involved. For set **28**, ρ is -0.85, but here it is a general narcosis which is involved rather than local skin anesthesia. The authors of this work, Buchi et al.,[69] measured the rate of hydrolysis of the compounds of set **28** in buffer at pH 11 and 50 °C, from which we formulated eq 19.

$$\log k = 1.55(\pm 0.09)\sigma + 1.79 \qquad n = 15,\ r^2 = 0.990 \quad \text{omitted: 4-OH} \tag{19}$$

For inhibition of hydrolysis, ρ would be -1.55, which is similar to ρ in set **22**. For the alkaline hydrolysis of X-$C_6H_4COOCH_3$, $\rho = 1.66$.[1] These two examples from physical organic chemistry can be compared to the enzymatic hydrolysis of X-$C_6H_4COOCH_2CH_2N^+Me_3$ by serum cholinesterase (eq 20).[178]

$$\log V_{max}/K_m = 1.55(\pm 0.48)\sigma + 0.50(\pm 0.32)E_S\text{-}2 - 13.5 \tag{20}$$

$$n = 13,\ r^2 = 0.879 \qquad \text{omitted: 2-F, 4-Me;}\ E_S\text{-}2 \text{ parameterizes ortho substituents}$$

In eq 20, ρ is similar to those mentioned above. Thus, it is highly likely that the favorable aspect of negative ρ in set **22** and **28** shows that electron-releasing groups inhibit hydrolysis, increasing the potency of the anesthetics; however, this does not rule out a binding role for the polarized carbonyl group as well.

A comparison of sets **22** and **28** can be made with set **24**. Inhibition of the maximum driving frequency is probably a narcotic-like action so that the log P' (P' is the distribution coefficient at pH 7.4) term is expected, but the negative ρ is unusual. It is conceivable that these compounds (set **24**) are interacting with the same receptor as the anesthetics (sets **22** and **28**) despite the fact that they are β-blockers and that this reaction is dependent on the electron density on the ether oxygen.

For the compounds of set **27**, the negative ρ might be related to stabilization of the Cl moiety from a side reaction of nucleophilic substitution. The effect of the two ring nitrogens and the two cyano groups would be to greatly activate Cl for nucleophilic attack.

Sets **25** and **29** correlate the inhibition of trypsin by benzamidines. The role of the $-\rho$ for substituents has been considered in a molecular modeling and QSAR analysis of trypsin inhibitors,[70] but only a tentative conclusion was reached. It may be that the increase in electron density on the aromatic ring increases its hydrophobicity[70] and thus strengthens binding to a hydrophobic pocket.

For sets **31**, **37**, and **43** the reason for the electronic effect of substituents on the dissociation of peroxide−cytochrome P-450 complexes is not obvious. However, it would be expected that a negative hydrophobic term, with $h = -1$, would favor the process. There are a number of examples for the binding of chemicals to P-450 with h near $+1$.[179]

An enormous amount of work has been done on the inhibition of photosystem II in chloroplasts. Camilleri et al. pointed out that by 1987 over 350 patents had been issued on this class of chemicals as herbicides, and many QSARs have been developed.[180]

Sets **35**, **36**, and **41** for the inhibition of photosystem II have small negative ρ's and Camilleri et al. present evidence that this is supported by other studies, but we have found that the majority of research on these inhibitors yield QSARs with very marginal ρ's. The QSAR depend largely on log P terms with a mean value of h near 1.[80]

Set **38** for the inhibition of *E. coli* by amino sulfones shows a typical bilinear relationship with hydrophobicity and a small negative ρ (-0.48). Both the amino group and the substituents would appear to increase toxicity by increasing the polarization of the sulfonyl group, but we have nothing with which to compare this set.

The enzymatic oxidation of phenylalanine derivatives (set **39**) is promoted by electron-releasing substituents. Presumably, this is associated with hydrogen abstraction from the α-carbon atom. The ρ of -0.46 seems reasonable. The only similar reaction we know of at present for comparison is the oxidation of X-$C_6H_4CH_2CH_2OH$ by chromic acid, which is correlated by eq 21.[181]

$$\log k = -0.99(\pm 0.26)\sigma - 0.06 \qquad n = 7,\ r^2 = 0.949 \tag{21}$$

The ρ's are suggestive despite the great disparity in reaction types.

We have nothing with which to compare set **40**.

The rather different phenylureas acting on chloroplasts (sets **36** and **41**) and mice (set **42**) have very similar ρ's (-0.56 and -0.44 vs -0.43). However, the values of h are positive for the plants and negative for the mice. The mouse system also contains steric terms. In the case of the chloroplasts reaction is with the thylakoid membrane, while the mice action may be with a membrane in the CNS. Comparative QSAR directs one's attention to quite unusual situations which may be worth investigation.

Two sets (**44** and **47**) correlate substituted anilines inhibiting photophosphorylation in chloroplasts and have similar ρ values and hydrophobic terms. The intercepts cannot be compared since the compounds in set **44** are partially ionized at pH 7.4, while those in set **47** are neutral. Positive ρ values indicate that lowering the electron density on the NH is beneficial, even in the case of $-C_6H_4NHC_6H_4-$. Set **47** can be

compared to set **49**. The two equations, including the intercepts and the quality of correlation, are very similar. It is possible that in both instances similar phosphorylation processes are involved.

The antimicrobial action of the phenols of set **50** can be compared with that of set **63**. Note especially the close agreement between the intercepts. The positive ρ suggests that increasing the polarity and hydrogen bonding character might increase toxicity. At first glance these sets appear to compare with the thiophenols of set **67**. However, the intercept in set **67** indicates that the thiophenols are roughly 1000 times more toxic. Obviously a different mechanism of action is involved.

The phenols in sets **95** and **108** have higher ρ's. Of particular interest is set **108**, which not only has a much higher ρ, but a much higher intercept. The intercepts cannot be compared as set **108** is based on π. We are unable to accurately calculate log P for this set of compounds. What the larger ρ portends is not evident.

A number of sets are based on the enzymatic hydrolysis of aryl glycines (hippurates): X-C_6H_4-$OCOCH_2NHCOC_6H_5$, sets **45**, **54**, **58**, **62**, **66**, and **70**; 3-pyridyl-$OCOCH_2NHCOC_6H_4$-X, set **46**; X-C_6H_4-$OCOCH_2NHSO_2Me$, sets **53**, **61**, and **64**; X-C_6H_4-$OCOCH_2NHR$, set **55** (R = SO_2Me or COC_6H_5); 4-HO-$C_6H_4CH_2CH(NHCOMe)CONHC_6H_4$-X, set **74**.

For these 12 examples, ρ varies from 0.39 to 0.78 with mean = 0.60 for six different hydrolases. In the examples of the sulfhydryl hydrolases, papain, actinidin, ficin, and the bromelains, k_{cat} is essentially constant, but in the case of chymotrypsin it is not. For those examples where k_{cat} is constant, it makes no difference if the the correlation is based on K_m or k_{cat}/K_m. What makes these sets of special interest is the low values of ρ. When esters of the type X-C_6H_4-$OCOCH_2NHCOC_6H_5$ are hydrolyzed in buffer at pH 6, $\rho = 1.91$, and at pH 8.0, $\rho^- = 1.66$. Thus we see that in the enzymic hydrolysis the enzyme needs much less help from the substituents compared to hydrolysis in buffer. As yet we have little experience with examples of this kind, so we cannot predict in advance when electronic effects on enzymic processes will parallel those for the corresponding nonenzymic cases. However, as we shall see in the following examples, ρ for enzymic substrates is normally lower than ρ for the corresponding nonenzymic process.

For the enzymatic hydrolysis of X-phenylacetals of sugars (sets **48**, **89**, and **97**), $\rho = 0.44$, 1.01, and 1.30. Why the value for set **48** is so much lower than the other two is not apparent. These results can be compared with eqs 22 and 23.

alkaline hydrolysis of X-C_6H_4-β-glucosides at 100 °C[182]

$$\log k = 2.47(\pm 0.36)\sigma^- - 2.97 \qquad n = 7,\ r^2 = 0.985 \tag{22}$$

alkaline hydrolysis of X-C_6H_4-deoxy-α-D-glucosides at 100 °C[183]

$$\log k = 2.47(\pm 0.73)\sigma^- - 4.55 \qquad n = 4,\ r^2 = 0.989 \tag{23}$$

The enzymes are obviously more effective in that they require less assistance from substituents (lower ρ) and operate at room temperature rather than at 100 °C. In the nonenzymatic reactions, σ^- gives significantly better correlations than σ, while in the enzymatic reactions, σ is better than σ^-.

For the present, there is nothing with which to compare sets **51** and **52**.

Set **56** seems to contain a surprise, since cellular absorption depends only on an electronic factor. One would expect to see a hydrophobic term. The authors conclude that the benzoylthiamines enter the cell by passive diffusion and then are hydrolyzed to the highly polar nondiffusable thiamine. Indeed, this is supported by the ρ of 0.56, which is similar to that for many examples of the enzymic hydrolysis of esters which we have considered.

The values of both ρ and h are very similar in sets **57** and **60**, suggesting similar substituent effects for quite different compounds acting on rather different enzymes. This would imply similarity of interaction sites.

Set **59** for LD_{30} of styrenes acting on *E. coli* illustrates rather nonspecific toxicity (notice the small intercept), mostly due to hydrophobic effects. However, the σ term does account for about 16% of the variance in the data.

Set **65** correlates the reduction in chlorophyll formation by phenylureas in radish plants. The positive ρ shows that the mechanism of action is different from phenylureas inhibiting photosystem II (sets **35**, **36**, and **41**). The two studies **36** and **41** have intercepts of 3.20, somewhat higher than the 2.40 of set **65**. That is, it takes a higher concentration of phenylurea to produce the bleaching action and the electronic effect of the substituents is quite different, showing that the phenylureas are acting in different ways. This would not be easy to see without a QSAR analysis.

Set **68** on the nonspecific toxicity of anilines resembles the QSARs for phenols (sets **50**, **63**, and **77**) in that there are positive σ and log P terms; however, the anilines seem to be intrinsically more toxic (compare intercepts). Set **68** can be compared with set **95**. The higher intercept with QSAR **95** again shows that isolipophilic anilines are more potent than phenols. (Since the coefficients are not identical, comparison of intercepts is a rough measure.)

The structure−activity relationships of simple phenols and anilines are not at all simple. Besides so-called nonspecific toxicity, which is largely log P dependent, both classes of compounds exhibit radical toxicity,[13] which is largely dependent on σ^+. The strongly ionized phenols uncouple oxidative phosphorylation. Careful QSAR studies enable one to distinguish the features which promote a particular type of toxicity.

Cholinesterase inhibitors of set **69** contain a positive ρ, the meaning of which is not yet clear. The negative E_S-3 term for meta substituents does relate to sets **130**, **133**, and **134** in Table 7. Since all substituents except H have negative E_S values in proportion to their steric effects, a negative coefficient with this term means a positive biological effect. In set **130**, the MR term is a measure of bulk, and thus in all three QSARs bulky 3-substituents increase

inhibitory potency when tested either on isolated enzyme or in whole flies (sets **130** and **133**).

Set **71** is based on a set of phenylureas somewhat different from those in set **65**; nevertheless, the ρ values are almost the same. Steric effects of the substituents are different in that for set **65** meta substituents produce a positive steric effect, while for set **71** para substituents produce a negative steric effect.

There are several sets of benzenesulfonamides: **72**, **73**, **80**, **84**, **86**, and **103**. For sets **72**−**86**, ρ ranges from 0.75 to 0.98, with a mean of 0.87. Four of these sets correlate inhibition of carbonic anhydrase as isolated enzyme, while the fifth (**72**) correlates diuretic activity in rats. For the whole animal there is a small negative log *P* term, while in the studies with the isolated enzyme all four sets have positive hydrophobic terms. The X-ray crystal structure of the enzyme is known, and QSAR/modeling studies show that substituents do contact hydrophobic space, but meta substituents also encounter steric effects.[184] However, hydrophilic compounds are better suited to reach the sites of action in the kidney. Thus, the advantage hydrophobic compounds have in binding to the enzyme is offset by the advantage hydrophilic compounds have in reaching the site of action. This is reflected in the QSAR of set **72**, where the hydrophobic term is of marginal value. Set **103** has a considerably higher ρ, which may be in part due to the narrow range in σ for the substituents examined. Also, in set **103** the action studied was binding to the enzyme, not K_i.

Kakeya et al.[109] measured the ionization constants for a set of benzenesulfonamides and obtained a ρ of 0.86. We have used their data and recently evaluated σ contants to obtain eq 24.

$$pK_a = 0.74(\pm 0.10)\sigma + 0.04 \qquad (24)$$

$$n = 14,\ r^2 = 0.956 \qquad \text{omitted: 4-NO}_2\text{, 4-NHCH}_3$$

The more ionized the sulfonamide group, the better it binds to a zinc atom in carbonic anhydrase.[184] The agreement of ρ of eq 24 with that in sets **72**, **73**, **80**, **84**, and **86** is reasonably good. Of course, activity in animals or cells will not linearly increase with increases in σ indefinitely. Too high a degree of ionization will inhibit the sulfonamides from crossing membranes.

Set **74** for the chymotrypsin hydrolysis of a set of amides has a ρ of 0.78, which is somewhat higher than that for the chymotrypsin hydrolysis of esters (set **117**).

Set **75** can be compared with set **85**. Although the ρ values for the two different kinds of acids are similar, the QSARs are different. The phenylacetic acid QSAR (set **75**) does not contain a term in the sterimol length parameter *L*, while this is a very important term is set **85**. In QSAR **75**, *L* is rather high, but this is also true of set **85**. The good agreement between ρ and *h* brings out the common features of the reaction mechanism, but it would appear that the extra atom in the phenoxyacetic acids introduces a steric problem requiring a steric parameter.

If the effect of substituents were associated with the carboxyl group, it would be expected that ρ would be larger for the phenylacetic acids. In fact, it is larger for the phenoxyacetics, which have one more atom between the ring and the carboxyl group. It has been postulated that the positive ρ is associated with a reaction in the plant with the benzene ring.[185]

The inhibition of photosystem II by the heterocycles of set **76** is distinctly different from that of the phenylureas of sets **35**, **36**, and **41**. In the former example, ρ is 0.81, while in the latter it is −0.57 to −0.44. All sets show an important hydrophobic term and all compounds probably interact with the thylakoid membrane, but the real meaning behind the electronic term is not obvious. This is likely to be associated with the basicity of the amino group.

It is unexpected that, in the influx of substituted phenylalanines into rabbit intestine (set **78**), hydrophobicity plays a negligible role and electronic effects dominate the QSAR. This would discount passive diffusion and indicate active transport where the acidity of the NH_2 or COOH might be important.

The role of σ in sets **79**, **82**, and **88** is obscure.

The electronic effect ($\rho = 0.90$) of substituents in set **81** is surprising, since the SO_3^- group is so highly ionized that one would not expect substituents to effect ionization. A search of the database turns up two QSARs of interest.

aqueous alkaline hydrolysis of X-$C_6H_4SO_3Me$ at 50 °C[186]

$$\log k = 0.92(\pm 0.03)\sigma - 3.71 \qquad n = 5,\ r^2 = 1 \quad (25)$$

ionization of X-$C_6H_4SO_3H$ in sulfuric acid[187]

$$\log k = 0.62(\pm 0.14)\sigma + 6.65(\pm 0.06)$$
$$n = 7,\ r^2 = 0.960 \quad (26)$$

The electronic effect in set **83** is the same as that of set **81**. The binding of the sulfate esters to the enzyme also appears to parallel the ability of the sulfur to react with a nucleophilic moiety.

In both eqs 25 and 26, we find modest positive values of ρ near that for set **81**. Lowering electron density on sulfur favors all three processes, but how this works with the erythrocytes is not clear.

A study of the acid hydrolysis of X-$C_6H_4OSO_3H$ in aqueous solution yields eq 27.[188]

$$\log k = 0.61(\pm 0.06)\sigma^- - 4.08(\pm 0.04)$$
$$n = 14,\ r^2 = 0.973,\ s = 0.52 \quad (27)$$

In this nonenzymatic reaction, σ^- is considerably better than σ ($r^2 = 0.723$).

In set **87** for the inhibition of fungi by phenyl guanidines, σ is the most important parameter. HB is assigned the value of 1 for several substituents capable of hydrogen bonding. However, we have nothing with which to compare this QSAR, nor do we have anything for comparison with set **90**.

In the example (set **91**) of the enzymic hydrolysis of what could be considered a set of benzoate esters, we do not find a hydrophobic term often seen with serum esterases. The value of ρ is significantly lower

than that for the hydrolysis of methyl benzoates (1.66), but not as low as that for some other esterases (sets **45**, **53**, **54**, **55**, **61**, **62**, **64**, **66**, **70**, and **74**). This liver esterase does not seem to be as effective as the others in bringing these esters into the transition state. Substituent help is welcome.

Sets **92** and **102** are for the inhibition of mitochondrial succinate dehydrogenase. The structures of these compounds are somewhat like those of set **8** and their analogs, which have been presumed to inhibit this enzyme. However, the ρ values are far apart: -1.93 vs 1.11 and 1.54. Thus, it seems unlikely that the action of the pesticides of set **8** is on this dehydrogenase.

It seems likely that the oxidation of compounds of set **93** occurs by a radical attack on the methylene moiety of $-OCH_2O-$. The clue to this possible mechanism comes from the fact that no hydrophobic term occurs in the QSAR of set **93**. Oxidation by cytochrome P-450 generally yields QSAR with hydrophobic terms.[189] Radical oxidation (by $^{\bullet}OH$) is notable in that the QSAR lack hydrophobic terms or have small negative h values.[13]

The enzymatic hydrolysis of phenyl acetates of set **94** show a good correlation with $\rho = 1.2$ for log $1/K_m$. For the correlation with log k_{cat}/K_m, ρ is almost the same, 1.09. The MR terms in two equations are also the same. Hydrolysis seems to be largely controlled by the K_m step. These results can be compared with eq 28.

alkaline hydrolysis of X-$C_6H_4OCOCH_3$ in aqueous 40% acetone at 15 °C[190]

$$\log k = 1.48(\pm 0.52)\sigma - 0.64 \qquad (28)$$

$$n = 6,\ r^2 = 0.970 \qquad \text{omitted: 4-COOH, 4-NH}_2$$

One might expect set **94** or eq 28 to be better correlated by σ^-, but σ is better in both instances. Again, we see enzymatic catalysis has a lower ρ than the nonenzymic process.

In set **96**, we have another example of a substituted phenoxy moiety which is correlated with a positive ρ (see sets **57**, **60**, and **85**). The reasons behind this are not clear.

Set **98** involves N_1-substituted phenyl sulfanilamides whose antibacterial activity depends on the inhibition of folate synthetase.[191] It is the ionized form of the sulfonamide group which is the active species; hence, a positive ρ is to be expected. The value of 1.31 is higher than that seen for the simple sulfonamides inhibiting carbonic anhydrase (sets **72**, **73**, **80**, **84**, and **86**). This may be the result of having the NH_2 conjugated with the $-SO_2NH-$ unit, which would decrease the acidity of $-SO_2NH-$.

The 5-X-8-OH-quinolines of set **99** can be regarded as a type of phenol; however, the effect of the substituents para to the OH group is complex. The positive ρ and h values are what is expected for the nonspecific toxicity of phenols, but the negative MR term shows that the phenols are not acting in the usual "nonspecific" fashion. The large intercept also is higher than that usually seen for simple phenols; however, P' indicates that P has been corrected for ionization, and this effects the intercept. Increasing the acidity of the phenols appears to increase their toxicity.

The cholinesterase hydrolysis of benzoate esters (set **100**) has a ρ value close to that for the alkaline hydrolysis of methylbenzoates (1.56) and a negative steric effect for ortho substituents (0.50 E_S-2). In this instance the enzyme does not show the usual low level of electronic effect of the substituents. One wonders if the charged group has an adverse effect on achieving the optimum transition-state structure.

Sets **101** and **104** have almost identical ρ's. In each case, the focus of the electronic effect of X could be on a benzylic H. Of course, the QSARs are quite different otherwise. It is conceivable that oxidation of these benzylic hydrogens could be involved. Comparative QSAR calls one's attention to similar situations of which one would not normally think.

The high positive value of ρ for set **105** suggests that it is the ionized form of the salicylic acids which is the inhibitor.

Sets **106** and **107** for the deacylation of X-benzoylchymotrypsins have ρ near that for the alkaline hydrolysis of methyl benzoates (1.66). There are no other terms in these QSARs which inply that the benzoyl moiety does not contact the enzyme. Under these conditions, one could expect deacylation to parallel hydrolysis. That is, the process does not appear to be enzyme mediated.

Set **109** for the inhibition of methyltransferase by benzoic acids has a large ρ but with such large confidence limits that one is not sure of its magnitude. Nevertheless, its sign implies that it is the anionic form which is important for inhibition. This is similar to set **105** and in contrast to sets **18** and **34**.

The large ρ (2.27) for the toxic action of isothiocyanates to *E. coli* in set **111** and the lack of a hydrophobic term make this set of special interest.[1]

$$\text{X-C}_6\text{H}_4\text{N=C=S} + \text{C}_2\text{H}_5\text{OH} \rightarrow \text{X-C}_6\text{H}_4\text{OC(=S)NH}_2$$

$$\log k = 2.16\sigma - 4.80 \qquad n = 8,\ r^2 = 0.951 \qquad (29)$$

$$\text{X-C}_6\text{H}_4\text{N=C=S} + \text{C}_6\text{H}_5\text{NH}_2 \rightarrow \text{X-C}_6\text{H}_4\text{NHC(=S)NHC}_6\text{H}_5$$

$$\log k = 2.14\sigma - 3.13 \qquad n = 4,\ r^2 = 0.988 \qquad (30)$$

The values of ρ in eqs 29 and 30 are essentially the same as that of set **111**. This suggests that almost any nucleophilic group in the bacteria would be a potential site for reaction.

In set **113**, the ρ (2.77) for hydrolysis of the X-benzoylpapain is much higher than for the chymotrypsins (sets **106** and **107**). Papain is a thiol esterase while chymotrypsin is a serine esterase. Cleavage of the C(=O)—S— bond is more sensitive to substituent effects than cleavage of C(=O)—O—.

We now turn to examples in Table 7, where correlation is dependent on σ^- rather than σ. As mentioned earlier, the quality of the correlation is often much the same, especially when a poor selection

of substituents has been made. However, when σ^- is significantly better, it does provide additional insight.

Set **114** correlates the antitumor activity of a set of platinum amines. It has been assumed that these compounds are effective by virtue of their ability to react with the DNA of the rapidly growing tumor cells more effectively than the normal cells. It has been noted that the negative ρ implies increased electron density on the amino groups and increases potency. This also would tend to strengthen the Pt−N bond.[13] This is in contrast to set **132**, where we find a positive ρ for mutagenesis and we are more certain that the reaction is with DNA. In the case of the antimelanoma activity (set **114**), other nucleophiles may be involved in the toxic action to tumor cells. This is reminiscent of the example of the aniline mustards, where animal data has a lower ρ than model systems or cell data. In neither set **114** nor **132** is there a *positive* hydrophobic effect, which is another example of electrophiles acting without benefit of hydrophobic interaction. It has been suggested that for set **132**, one of the amino groups acts as the leaving group.[13] The important question is why would Cl function in one instance as the leaving group and NH_2 do so under other circumstances? This may be related to the geometry of the site of interaction and the positioning of platinum amine in it.

Set **115** is based on the same compounds as set **69** for the inhibition of acetylcholinesterase, but the QSARs are quite different. For set **69**, ρ is positive (0.73), while in set **115** it is negative. How this is to be explained remains to be elucidated. This situation is similar to that of set **114** and **132**. The same compounds give quite different results under different circumstances.

Set **116** has a negative ρ, indicating that an increase in the electron density on oxygen promotes the catalytic step in hydrolysis. The reaction is complex. The overall reaction k_{cat}/K_m is best correlated by an equation parabolic in σ.

Set **117** for the chymotrypsin hydrolysis of phenyl hippurates has a ρ similar to those for the papain, ficin, actimidin, and bromelain hydrolysis of these esters, except that with chymotrypsin, σ^- gives a distinctly better correlation than σ. Since there is a good selection of substituents in most of these examples, this appears to be transition state related. The thiol hydrolases behave differently from the serine hydrolase (set **117**).

In set **118**, inhibition of hydrolysis of alkaline phosphatase by X-$C_6H_4OPO_3H$ shows a small dependence on σ^- (ρ = 0.49). The electronic effect is somewhat like that of set **83**, but in **83** hydrolysis occurs and in **118** inhibition results. It is possible that nucleophilic substitution occurs with the phenoxide moiety acting as the leaving group. The small ρ is reminiscent of sets **45**, **46**, **53**−**55**, **61**, **62**, **64**, **66**, and **70**.

In the inhibition of *S. fecalis* by 2,4-(X-$C_6H_4NH)_2$-pyrimidines (set **119**), σ^- is more important than σ. Delocalization of the lone pair electrons on NH is no doubt involved, but its mechanistic significance is not clear. The effect is similar to that seen in sets **47**, **49**, and **68**.

In the conjugation of phenols with sulfate (set **120**), a rather complex QSAR is found with negative steric effects in the meta and para positions. Electron withdrawal by substituents plays a small part, suggesting that increased polarity, possibly via hydrogen bonding, plays a role in complex formation.

In set **121**, ρ is relatively small as is the case for other hydrolases acting on this type of ester.

Set **122** is similar to set **98**, where the same compounds are acting on *E. coli*. With the bacteria, σ is slightly better than σ^-; however, with enzyme inhibition (presumably the root cause of bacterial toxicity), σ^- is significantly better than σ.[191]

In set **123**, the ρ of 0.87 shows that increasing the potential for ionization increases toxicity. However, this value of ρ is far from that of >2 found for the ionization of phenols.

The QSAR of set **124** is not very good (too few data points), but it is of interest when compared with somewhat different guanidines (set **87**) inhibiting fungi (ρ = 0.98). Building up lateral support from other QSAR can strengthen the value of small sets.

Set **125** correlates the enzymic hydrolysis of phenylacetates. The result is similar to set **94** (ρ = 1.09). Using σ in set **125** yields a ρ of 1.3 and r^2 of 0.837. This must be attributed to different transition states in the two different enzymes.

In the case of the correlation of hydrolysis of phenylacetates by carbonic anhydrase (set **110**), ρ is much higher. This enzyme seems less efficient than the other hydrolases (i.e. it needs more help from the substituents).

Set **127** for the inhibition of yeast by salicylaldehydes is similar to set **105** for the inhibition of glutamic acid dehydrogenase by salicylic acid, except that for the acids σ works better than σ^-. In both sets the variation in substituents is poor with respect to σ^-.

Starting with set **128**, there are a number of QSARs (**128**, **129**, **130**, and **133**−**138**) which are associated with the inhibition of insect cholinesterase or the toxic action to insects. It is interesting that in these examples σ^- is significantly better than σ. In the cleavage of the C_6H_5O-P or C_6H_5S-P (set **126**), delocalization of lone pair electrons on O or S plays (unlike the phenylacetates) an important role in the reaction with the isolated enzyme as well as in the examples where the endpoint is the LD_{50} of insects.

Indeed, the similarity of ρ for the isolated enzyme and the insecticidal activity is taken as evidence that it is the inhibition of cholinesterase in the insect that is responsible for the toxic action. For inhibition of the isolated enzyme, ρ ranges from 2.29 to 2.94, with a mean of 2.56. For the LD_{50} of insects, ρ ranges from 1.84 to 2.56, with a mean of 2.16. In none of these equations, whether isolated enzyme or whole insect, is there a positive hydrophobic term. This is quite different from vertebrate enzymes, where hydrophobicity is the dominant factor.[192]

Another feature of these QSARs which helps make mechanistic connection between the *in vitro* enzymic data and the *in vivo* insect toxicity is the steric effect of meta substituents. In sets **133**, **134**, and **138**, negative E_S terms show that meta substituents increase inhibitory potency (recall that E_S values are

scaled to H = 0 and are all negative). In set **130**, the bulk parameter MR provides a slightly better result than E_S. In set **128**, the indicator variable $I = 1$ for meta substituents yields a better equation than E_S or MR. This same parameter yields results equivalent to E_S or MR in **130**, **133**, and **138**. In set **134**, E_S is significantly better than MR or I.

The single example in which no steric effect could be detected for meta substituents is set **136** for the LD_{50} of thrips. This implies a receptor different from that of flies for this organism. In fact, the lack of close agreement with the coefficients for E_S suggest a variable portion in the enzyme. There is more than one form of the enzyme in flies.

Unfortunately, we have only one modest example for the hydrolysis of X-$C_6H_4P(=O)(OEt)_2$ at pH 7.6, 37 °C, which can be compared with the action of these esters on cholinesterase and flies (eq 31).[193]

$$\log k = 2.94(\pm 0.78)\sigma - 6.81$$
$$n = 4,\ r^2 = 0.993,\ s = 0.118 \quad (31)$$

Using σ^- in eq 31 gives an inferior result ($r^2 = 0.940$). As usual, the ρ values for the enzymic reactions are, with one exception, smaller than that of eq 31. However, the differences are not large suggesting that ligand binding does not appear to favor enzymic efficiency.

The QSARs for acetylcholinesterase and flies can be compared with eq 32 for the alkaline hydrolysis of X-$C_6H_4OP(=O)(OC_2H_5)_2$.[194]

$$\log k = 1.25(\pm 0.14)\sigma^- - 1.56(\pm 0.07) \quad (32)$$
$$n = 25,\ r^2 = 0.938,\ s = 0.151$$
omit: 2-Br and 2-Cl

ρ for eq 32 agrees with that found by Metcalf and Fukuto (1.19 ± 0.14).[195]

The more potent nucleophilic OH^- in the alkaline hydrolysis of eq 32 results in a much lower absolute value of ρ.

While both of examples for alkaline aqueous hydrolysis are best correlated by σ^-, the magnitude of ρ^- is much less than for the biological reactions. In these examples substituent assistance is much more important than in aqueous hydrolysis. However, we have more to learn from examples where ρ values from biological and physical organic correlations are in qualitative, but not quantitative agreement. No doubt this is related to differences in the transition states, but exactly how is not clear.

Discussion

In this review we have considered a relatively small set of biological QSARs of reasonable quality which contain terms in σ or σ^-. A problem which has nagged us from the beginning of the QSAR paradigm is to what degree can we expect electronic effects defined by Hammett constants for heterogenous biological systems to resemble those for comparable organic reactions in homogenous solutions. This is a difficult problem in part because the biological tests, especially in whole organisms, cannot be made with the same high precision that is common in the measurements of reaction equilibria or rate studies of physical organic chemistry. Confounding the problem is the heterogeneity of enzymes and receptors and, in living systems, the many possible metabolic side reactions. The results in Tables 6 and 7 show that our assumption is reasonable.

Another serious difficulty stems from the way those who make synthetic variations of lead compounds for biological testing make their choices. Rather few synthetic organic chemists have any background in QSAR and thus lack understanding of the collinearity problem. Moreover, they are often handicapped by the difficulty in synthesis of the complex molecules which are often their starting points for developing a set of congeners. Electronic effects need to be uncovered by relatively small carefully selected substituents to mitigate possible steric effects and in such a way that collinearity with hydrophobic effects is limited. Without such attention to detail, one cannot gain a clear view of the role of substituent effects.

Since most discoveries of new drugs have been accidental, there is still a strong tendency to play hunches early on in drug design. Making wild structural changes should come after some feeling for the relative importance of steric, electronic, and hydrophobic changes has been obtained via conservative structural modification. Unless good variation is present in the properties of the substituents, inference cannot be made with confidence.

Still, we have much to learn about intra- and intermolecular steric effects. It is a negelected area. Physical organic chemists have not rushed to study even the effects of ortho substituents since Taft first defined E_S. Their primary interest has been to use a minimum set of "well-behaved" substituents to define a value of ρ. Those designing bioactive compounds are often playing molecular roulette (now on a grand scale with combinatorial synthesis) by prospecting with unusual structural changes not easily parameterized. Rational drug design has come to mean the use of molecular graphics which tends to consider mostly steric properties. Published studies in which well-designed congener sets have been tested on simple organic reactions, with receptor or enzyme, in cells, and finally in whole multicellular organisms, are almost nonexistant. The aniline mustards are the best example of our ability to correlate electronic effects from QSAR in animals to QSAR in cells and then to QSAR from physical organic chemistry. Possibly such results do exist in the files of the drug or agrochemical companies, but are not available to the public.

Thus, we must make do with what is available, and this is difficult because the needed publications are scattered throughout the literature. This has provided the motivation for the construction of our QSAR database. The best data for comparative QSAR in our present study are the mustards, the insecticides at the end of Table 7, and the carbonic anhydrase inhibitors (sets **72**, **80**, **84**, **86**, and **103**).

In the case of the mustards acting in animals to improve survival after being challenged with leukemia, the results are far better than we would have dreamed possible before QSAR got well underway.

The great toxicity of the mustards to the host, their propensity to react with all nucleophiles, even those as weak as water, would discourage most from even attempting QSAR studies. However, as noted in the introduction, QSAR studies with whole organisms are not grossly inferior to those with enzymes. The expense of such studies, their interdisciplinary nature, and the obvious multivariate complexity of the heterogenous reaction systems have, unfortunately, discouraged most theoretically inclined academicians from entering the fray.

Of course, more must be learned about metabolism so that, to some degree, these problems can be sidestepped.[189] At present, one hopes that metabolic problems do not preclude getting a meaningful QSAR. In the study of insecticides, metabolic inhibitors have sometimes been used to inhibit these side reactions with some success. Avoiding metabolically sensitive substituents such as esters or nitro compounds is helpful. Esters are vulnerable to a variety of esterases and should be avoided until one has some notion of the shape of the QSAR. The aromatic nitro group, often attractive to synthetic chemists, is easily reduced by cytosolic reductases to the toxic hydroxylamines. Hydrophobic compounds are in general more rapidly attacked by the P-450 enzymes than hydrophilic substances. This needs to be considered in interpreting the character of hydrophobic terms in QSAR, etc. Despite all of these caveats, we believe that the results in Tables 6 and 7 clearly demonstrate that there is much to be gained from the systematic application of what has been learned from mechanistic organic and biochemistry to the problems of drug design and toxicology.

A generalization which seems to be emerging is that ρ for enzymic reactions with substrates is smaller in magnitude than that for the corresponding nonenzymatic chemical reaction. This might well have been anticipated, but we were slow to appreciate this point. The virtue of enzymes, as Pauling pointed out years ago, is that they can constrain the substrate in a conformation near to that of the transition state. Thus, they can dispense with much of the electronic help of substituents. In fact, it may be possible to categorize enzyme efficiency by comparison of ρ values.

Another insight about ligand interactions with enzymes (or receptors in general) is that steric effects are not "all or none" as implied by the lock and key concept. The fact that empirical parameters such as E_s, MR, and the sterimol parameters can account, usually in a linear fashion, for steric effects has important implications for QSAR in general, no matter which approach is used. From graphics-QSAR analysis,[6] it seems clear that there is no fixed position in which a parent molecule and its simple derivatives fit to a receptor site. Activity falls off linearly (often over considerable range) as larger substituents are introduced at a given position. Very likely there is some "give" by the receptor surface as well as movement by the ligand from the ideal binding position. Although correlation by means of an empirical parameter in traditional QSAR says nothing about where most of the movement (enzyme or ligand) occurs, our studies[6] lead us to believe that

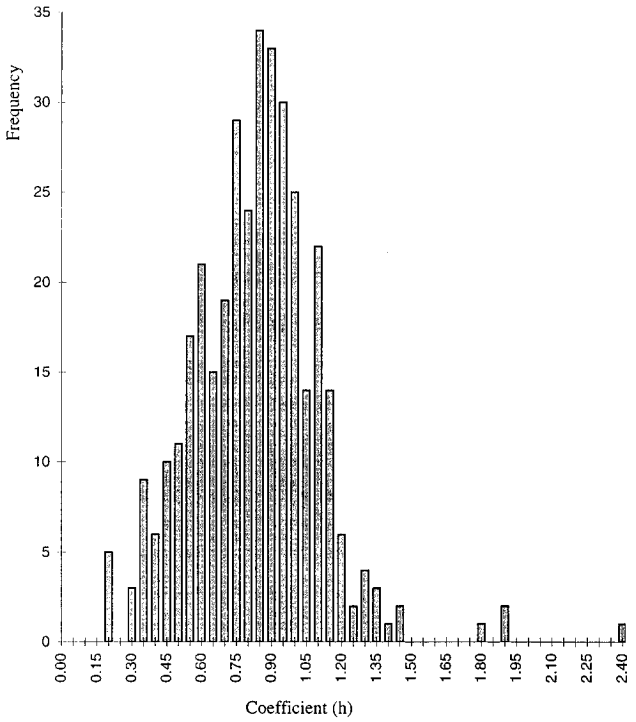


Figure 1. The distribution of the coefficient with log P (h) for all examples (363) for equations containing only a single log P term for data sets having five or more compounds with $r > 0.95$.

it is mostly the ligand which yields from its ideal position. This implies that there is in fact a continuum of positions assumed by a set of congeners which are empirically accounted for by steric parameters. This is a significant problem for QSAR approaches, such as CoMFA, which attempt to define a fixed position for a set of congeners. This problem must be approached by the empirical methods of traditional QSAR and molecular docking methods.

Possibly the most important use of our current database which we have not directly considered in this report is for gaining perspective about a newly derived equation. The database can be searched in a few minutes to find all examples which have similar shape. A range can be placed on the coefficients of the various terms to see if similar examples have been reported in any field (ionization or chemicals affecting insects) or in the whole databank. As the databank grows, this will become extremely valuable.

The overall conclusion from the results in Tables 6 and 7 is that we are now able to see some consistency in the coefficients in the biological QSAR. There are parallels between physical organic and biological QSAR, although they are not as sharp as one might wish. This is to be expected, considering the complexity of understanding how a set of organic chemicals interacts with the innumerable macromolecules, receptors, and membranes in an entity even as simple as a cell. Nevertheless, we see order emerging. From our knowledge of the parameters of organic reactions obtained from physical organic chemistry, we can begin to have expectations when studying many types of chemicals with enzymes, receptors, and cells. This is also true for hydrophobic interactions, as illustrated with Figures 1 and 2.

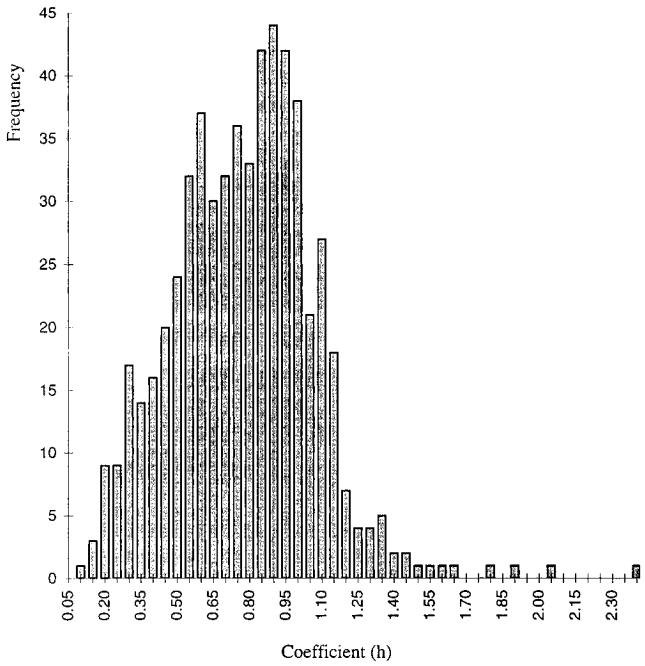


Figure 2. The distribution of the coefficient with log *P* for all examples (578) for equations containing a log *P* term as well as for one or more other terms for data sets having five or more compounds with $r > 0.95$.

In the histogram of Figure 1 for the simplest of equations (only a linear term in log *P*) of reasonable quality, most of the coefficients (*h*) (90%) fall in the range 0.45−1.25 and 40% are in the range 0.80−1.00. We have shown only those examples with positive *h* values, since there are as yet relatively few with negative values. There are two factors which determine *h*: movement of the chemical through hydrophobic and hydrophilic compartments and then partitioning onto the ultimate site for triggering a biological response. The value of *h* will be set by these two processes. At present there are not enough good data to delineate the relative importance of these. In Figure 2, the 578 examples are shown for QSAR which contain other terms in addition to log *P*. Again the highest frequency centers in the region where *h* is 0.8−0.9. The gross distribution in Figure 2 is similar to Figure 1. Ninety percent of the coefficients fall in the range 0.3−1.2 and 41% are in the range 0.75−1.00.

The histogram of Figure 2 is somewhat skewed toward the lower values of *h*. The reason behind this will require further study, but the problem may lie in getting a sharper separation of steric and hydrophobic properties of the set members. This, of course, goes back to the design of the congeneric sets. Nevertheless, we now have expectations of what constitutes "normal" hydrophobic effects in QSAR.

Doing comparative studies such as those in Figures 1 and 2 as well as those in Tables 6 and 7 calls our attention to consistencies or inconsistencies, which can spark new investigations. For example, in Figures 1 and 2, there are three spikes at h of 0.60, 0.75, and 1.10 which need to be studied.

The understanding of QSAR suffers from many of the same problems as organic chemistry itself. We all use equations for general reactions only to discover later that the generalization has very significant limitations. Frequently the synthetic chemist sees the yields of the desired product dropping to a few percent due to side reactions in a way reminiscent of metabolic side reactions. Nevertheless, organic chemists have persisted to develop a truly great science. Although the problems for biological QSAR are more formidable, we are optimistic that it too can be developed into a science, with the intellectual strength and practical utility of synthetic organic chemistry, to organize our understanding of how organic compounds affect the many forms of life and its component parts.

Acknowledgments. This work was supported in part from grant 1 R01 ES 07595-01 from NIEH and in part from a grant for fundamental studies in toxicology from R. J. Reynolds. We thank Toshio Fujita, Ki Kim, Albert Leo, Eric Lien, and Maurizio Recanatini for many helpful suggestions in preparing this review.

References

(1) Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147.
(2) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology;* American Chemical Society: Washington, DC, 1995.
(3) Wold, S.; Johansson, E.; Cocchi, M. In *3D QSAR* in *Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993, p 523.
(4) Cramer, R. D., III; DePriest, S. A.; Patterson, D. E.; Hecht, P. In *3D QSAR* in *Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993, p 443.
(5) *Comprehensive Medicinal Chemistry*; Ramsden, C. A., Ed.; Pergamon: Tarrytown, NY, 1990; Vol. 4.
(6) Hansch, C.; Klein, T. *Acc. Chem. Res.* **1986**, *19*, 392.
(7) Hansch, C.; Kim, D.; Leo, A. J.; Novellino, E.; Silipo, C.; Vittoria, A. *Critical Rev. Toxicol.* **1989**, *19*, 185.
(8) Hansch, C.; Leo, A.; Elkins, D. *J. Chem. Doc.* **1974**, *14*, 57.
(9) Elkins, D.; Leo, A.; Hansch, C. *J. Chem. Doc.* **1974**, *14*, 65.
(10) Jaffe, H. H. *Chem. Rev.* **1953**, *53*, 191.
(11) Hoekman, D.; Leo, A.; Li, P.; Zhang, L.; Hansch, C. *Biotechnology Computing.* In *Proceedings of the Twenty-Seventh Annual International Conference on Systems Science.* Hunter, L., Ed.; IEEE Computer Society Press: Piscataway, NJ, 1994; Vol. V, 193.
(12) Hansch, C.; Hoekman, D.; Leo, A.; Zhang, L.; Li, P. *Toxicol. Lett.* **1995**, *79*, 45.
(13) Hansch, C.; Telzer, B. R.; Zhang, L. *Crit. Rev. Toxicol.* **1995**, *25*, 67.
(14) Weininger, D.; Weininger, J. L. In *Comprehensive Medicinal Chemistry*; Ramsden, C. A., Ed.; Pergamon Press: Tarrytown, NY, 1990; Vol. 4, 59.
(15) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.
(16) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology;* American Chemical Society: Washington, DC, 1995; Chapter 2.
(17) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants;* American Chemical Society: Washington, DC, 1995.
(18) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology;* Wiley Interscience: New York, 1979. (a) Krygowski, T. M.; Wrona, P. K.; Zielkowska, V.; Reichardt, C. *Tetrahedron* **1985**, *41*, *4519*. (b) Reichardt, C.; Harbugch-Görnert. *Liebigs. Ann. Chem.* **1983**, 721.
(19) Wells, P. R. *Linear Free Energy Relationships*; Academic Press: New York, 1968; p 25.
(20) Shorter, J. *Correlation Analysis in Organic Chemistry: an Introduction to Linear Free-Energy Relationships*; Clarendon Press: Oxford, 1973; p 1.
(21) Ishida, S.; Ida, M. *Agric. Biol. Chem.* **1967**, *31*, 410.
(22) Frederick, J.; Dippy, F. J.; Williams, F. R. *J. Chem. Soc.* **1934**, 161.
(23) Trachtenberg, E. N.; Odian, G. *J. Am. Chem. Soc.* **1958**, *80*, 4018.
(24) Frederick, J.; Dippy, F. J.; Page, J. E. *J. Chem. Soc.* **1938**, 357.
(25) Hayes, N. V.; Branch, G. E. K. *J. Am. Chem. Soc.* **1943**, *65*, 1555.
(26) Pasto, D. J.; McMillan, D.; Murphy, T. *J. Org. Chem.* **1965**, *30*, 2688.
(27) Pettit, L. D.; Royston, A.; Sherrington, C.; Whewell, R. J. *J. Chem. Soc. B.* **1968**, 588.
(28) Roberts, J. D.; McElhill, E. A. *J. Am. Chem. Soc.* **1950**, *72*, 628.
(29) Roberts, J. D.; McElhill, E. A.; Armstrong, R. *J. Am. Chem. Soc.* **1949**, *71*, 2923.

(30) Fuchs, R.; Kaplan, C. A.; Bloomfield, J. J.; Hatch, L. F. *J. Org. Chem.* **1962**, *27*, 733.
(31) Standoli, L. *J. Chem. Soc., Perkin Trans. 2* **1975**, 371.
(32) Fuchs, R.; Bloomfield, J. J. *J. Org. Chem.* **1966**, *31*, 3423.
(33) Benghiat, I.; Becker, E. I. *J. Org. Chem.* **1958**, *23*, 885.
(34) Molho, D.; Giraud, M. *Bull. Soc. Chim. Fr.* **1969**, 4447.
(35) Hogeveen, H. *Recl. Trav. Chim. Pays-Bas.* **1964**, *83*, 813. (a) Fadnavis, N. W.; Bagavant, G. *Indian J. Chem.* **1979**, *17B*, 518.
(36) Bowden, K.; Chapman, N. B.; Shorter, J. *Can. J. Chem.* **1964**, *42*, 1979.
(37) Kindler, V. K. *Justus Liebigs Ann. Chem.* **1926**, *450*, 1.
(38) Kindler, V. K. *Justus Liebigs Ann. Chem.* **1927**, *452*, 90.
(39) Kindler, V. K. *Justus Liebigs Ann. Chem.* **1928**, *464*, 278.
(40) Fuchs, R. *J. Org. Chem.* **1963**, *28*, 3209.
(41) Evans, D. P.; Gordon, J. J.; Watson, H. B. *J. Chem. Soc.* **1937**, 1430.
(42) Fuchs, R.; Caputo, J. A. *J. Org. Chem.* **1966**, *31*, 1524.
(43) Jones, B.; Watkinson, J. G. *J. Chem. Soc.* **1958**, 4064.
(44) Bloomfield, J. J.; Fuchs, R. *J. Org. Chem.* **1961**, *26*, 2991. (a) Eaborn, C.; Eastmond, R.; Walton, D. R. M. *J. Chem. Soc. B* **1970**, 752.
(45) Kim, C. J.; Brown, H. C. *J. Am. Chem. Soc.* **1969**, *91*, 4289. (a) Ouellette, R. J.; Papa, R.; Attea, M.; Levin, C. *J. Am. Chem. Soc.* **1970**, *92*, 4893.
(46) Hansch, C.; Björkroth, J. P.; Leo, A. *J. Pharm. Sci.* **1987**, *76*, 663.
(47) Everett, J. L.; Roberts, J. J.; Ross, W. C. J. *J. Chem. Soc.* **1953**, 2386.
(48) Palmer, B. D.; Wilson, W. R.; Pullen, S. M.; Denny, W. A. *J. Med. Chem.* **1990**, *33*, 112.
(49) Denny, W. A.; Wilson, W. R.; Palmer, B. D. In *Probing Bioactive Mechanism*; American Chemical Society: Washington, DC, 1989; p 291.
(50) Gourdie, T. A.; Valu, K. K.; Garvatt, G. L.; Bortizki, T. J.; Baguley, B. C.; Wakelin, P. G.; Wilson, W. R.; Woodgate, P. D.; Denny, W. A. *J. Med. Chem.* **1990**, *33*, 1177.
(51) Andres, H. H.; Vogel, R. S.; Tarr, G. E.; Johnson, L.; Weber, W. W. *Mol. Pharmacol.* **1987**, *31*, 446.
(52) Equation derived from unpublished data from the National Cancer Institute.
(53) Panthananickal, A.; Hansch, C.; Leo, A.; Quinn, F. R. *J. Med. Chem.* **1978**, *21*, 16.
(54) Snel, M.; von Schmeling, B.; Edington, L. V. *Phytopathology* **1970**, *60*, 1164.
(55) Cornell, N. W.; Hansch, C.; Kim, K. H.; Henegar, K. *Arch. Biochem. Biophys.* **1983**, *227*, 81.
(56) Tolf, B. R.; Dahlbom, R.; Akeson, A.; Theorell, H. *Acta Pharm. Suec.* **1985**, *22*, 147.
(57) Fiedler, H.; Weuffen, W. *Pharmazie* **1966**, *21*, 765.
(58) Panthananickal, A.; Hansch, C.; Leo, A. *J. Med. Chem.* **1979**, *22*, 1267.
(59) Fersht, A. R.; Blow, D. M.; Fastrez, J. *J. Biochem.* **1973**, *12*, 2035.
(60) Hayashi, H.; Penniston, J. T. *Arch. Biochem. Biophys.* **1973**, *159*, 563.
(61) Teng, N.; Kotowycz, G.; Clavin, M.; Hosokawa, G. *J. Biol. Chem.* **1971**, *246*, 5448.
(62) Ito, S.; Kojima, Y.; Fujimori, K.; Matsunari, K.; Shimazaki, I.; Suda, Y. *J. Pestic. Sci.* **1986**, *11*, 579.
(63) Ishida, S. *Agric. Biol. Chem.* **1966**, *30*, 800.
(64) Galinsky, A. M.; Gearien, J. E.; Perkins, A. J.; Susina, S. V. *J. Med. Chem.* **1963**, *6*, 320.
(65) Rauls, D. O.; Baker, J. K. *J. Med. Chem.* **1979**, *22*, 81.
(66) Andrews, J. M.; Roman, D. P.; Bing, D. H. *J. Med. Chem.* **1978**, *21*, 1202.
(67) Lowe, G.; Yuthavong, Y. *Biochem. J.* **1971**, *124*, 117.
(68) Nakamura, A.; Ikeda, O.; Segawa, H.; Takeuchi, Y.; Takematsu, T. *Agric. Biol. Chem.* **1983**, *47*, 2923.
(69) Buchi, J.; Bruhin, H. K.; Perlia, X. *Arzneim.-Forsch.* **1971**, *21*, 1003.
(70) Recanatini, M.; Klein, T.; Yang, C. Z.; McClarin, J.; Langridge, R.; Hansch, C. *Mol. Pharmacol.* **1986**, *29*, 436.
(71) Blake, R. C.; Coon, M. J. *J. Biol. Chem.* **1980**, *255*, 4100.
(72) Leo, A.; Panthananickal, A.; Hansch, C.; Theiss, J.; Shimkin, M.; Andrews, A. W. *J. Med. Chem.* **1981**, *24*, 859.
(73) Ohta, H.; Jikihara, T.; Wakabayashi, K.; Fujita, T. *Pestic. Biochem. Physiol.* **1980**, *14*, 53.
(74) Kasuya, F.; Igarashi, K.; Fukui, M. *J. Pharmacobio-Dyn.* **1991**, 14, 671.
(75) Mitsutake, K.; Iwamura, H.; Shimuzu, R.; Fujita, T. *J. Agric. Food Chem.* **1986**, *34*, 725.
(76) Takemoto, I.; Yoshida, R.; Sumida, S.; Kamoshita, K. *J. Pestic. Sci.* **1984**, *9*, 517.
(77) Coates, E. A.; Cordes, E. H.; Kulkarni, V. M.; Richter, M.; Schaper, K. J.; Wiese, M.; Seydel, J. K. *Quant. Struct.-Act. Relat.* **1985**, *4*, 99.
(78) Neims, A. H.; Deluca, D. C.; Hellerman, L. *Biochemistry* **1966**, *5*, 203.
(79) Claesen, M.; van Dijck, P.; Vanderhaeghe, H. *J. Pharm. Pharmacol.* **1954**, *6*, 127.
(80) Kakkis, E.; Palmire, V. C.; Strong, C. D.; Bertsch, W.; Hansch, C.; Schirmer, U. *J. Agric. Food Chem.* **1984**, *32*, 133.
(81) Pavia, R. M.; Lobbestael, J. S.; Taylor, P. G.; Hershenson, F. M.; Miskell, D. L. *J. Med. Chem.* **1990**, *33*, 854.
(82) van den Berg, G.; Brandse, M. *Z. Naturforsch.* **1984**, *39C*, 107.
(83) Carotti, A.; Raguseo, G.; Hansch, C. *Quant. Struct.-Act. Relat.* **1985**, *4*, 145.
(84) Compadre, C. M.; Hansch, C.; Klein, T. E.; Petridou-Fischer, J.; Selassie, C. D.; Smith, R. N.; Steinmetz, W.; Yang, C-Z.; Yang, G-Z. *Biochim. Biophys. Acta.* **1991**, *1079*, 43.
(85) Oettmeier, W. *Z. Naturforsch.* **1979**, *34C*, 1024.
(86) Tanaka, M; Kyosaka, S.; Sekiguchi, Y. *Chem. Pharm. Bull.* **1976**, *24*, 3144.
(87) Reisner, D. B.; Borick, P. M. *J. Pharm. Sci.* **1955**, *44*, 148.
(88) Weuffen, W.; Richter, L. *Pharmazie* **1970**, *25*, 480.
(89) Pratesi, P.; Grana, E.; Villa, L. *Proc. 3rd Int. Pharmacol. Meeting* **1968**, *7*, 283.
(90) Bunting, J. W.; Murphy, J.; Myers, C. D.; Cross, G. G. *Can. J. Chem.* **1974**, *52*, 2648.
(91) Carotti, A.; Smith, R. N.; Wong, S.; Hansch, C.; Blaney, J. M.; Langridge, R. *Arch. Biochem. Biophys.* **1984**, *229*, 112.
(92) Smith, R. N.; Hansch, C.; Kim, K. H.; Omiya, B.; Fukumura, G.; Selassie, C. D.; Jow, P. Y. C.; Blaney, J. M.; Langridge, R. *Arch. Biochem. Biophys.* **1982**, *215*, 319.
(93) Hansch, C.; Calef, D. F. *J. Org. Chem.* **1976**, *41*, 1240.
(94) Shindo, H.; Okamoto, K.; Totsu, J. *Chem. Pharm. Bull.* **1967**, *15*, 295.
(95) Hansch, C.; Coats, E. *J. Pharm. Sci.* **1970**, *59*, 731.
(96) Carotti, A.; Casini, G.; Hansch, C. *J. Med. Chem.* **1984**, *27*, 1427.
(97) Suguira, K.; Goto, M. *Chem.-Biol. Interact.* **1981**, *35*, 71.
(98) Hulbert, P. B. *Mol. Pharmacol.* **1974**, *10*, 315.
(99) Carotti, A.; Raguseo, C.; Hansch, C. *Chem.-Biol. Interact.* **1985**, *52*, 279.
(100) Hawkins, H. C.; Wiliams, A. *J. Chem. Soc., Perkin Trans. 2* **1976**, 723.
(101) Hansch, C.; Lien, E. J. *J. Med. Chem.* **1971**, *14*, 653.
(102) Takemoto, I.; Yoshida, R.; Sumida, S.; Kamoshita, K. *Pestic. Biochem. Physiol.* **1985**, *23*, 341.
(103) Ballio, A.; Cingolani, E. *E. Boll. Soc. Ital. Biol. Sper.* **1953**, *29*, 622.
(104) Arnold, L. M.; Lin, D. T.; Schultz, T. W. *Chemosphere* **1990**, *21*, 183.
(105) Cavallito, C. J.; Yun, H. S.; Kaplan, T.; Smith, J. C.; Foldes, F. F. *J. Med. Chem.* **1970**, *13*, 221.
(106) Carotti, A.; Hansch, C.; Mueller, M. M.; Blaney, J. M. *J. Med. Chem.* **1984**, *27*, 1401.
(107) Nakagawa, Y.; Kitahara, K.; Nishioka, T.; Iwamura, H.; Fujita, T. *Pestic. Biochem. Physiol.* **1984**, *21*, 309.
(108) Kakeya, N.; Yata, N.; Kamada, A.; Aoki, M. *Chem. Pharm. Bull.* **1970**, *18*, 191.
(109) Kakeya, N.; Yata, N.; Kamada, A.; Aoki, M. *Chem. Pharm. Bull.* **1969**, *17*, 2558.
(110) Phillip, M.; Pollack, R. M.; Bender, M. L. *Proc. Nat. Acad. Sci. U.S.A.* **1973**, *70*, 517.
(111) Muir, R. M.; Fujita, T.; Hansch, C. *Plant Physiol.* **1967**, *42*, 1519.
(112) Cammilleri, P.; Astles, D. P.; Kerr, M. W.; Spencer, J. E. *J. Agric. Food Chem.* **1990**, *38*, 1601.
(113) Jansson, K.; Jansson, V. *Toxicol. Lett.* **1993**, *69*, 289.
(114) Hajjar, J. J.; Curran, P. F. *J. Gen. Physiol.* **1970**, *56*, 673.
(115) Selassie, C. D.; Hansch, C.; Khwaja, T. A.; Dias, C. B.; Pentecost, S. *J. Med. Chem.* **1984**, *27*, 347.
(116) Lien, E. J.; Hussain, M.; Tong, G. L. *J. Pharm. Sci.* **1970**, *59*, 865.
(117) Cabantchik, Z. I.; Knauf, P. A.; Rothstein, A. *Biochem. Biophys. Acta.* **1978**, *515*, 239.
(118) Guo, Z. R.; Dietrich, S. W.; Hansch, C.; Dolinck, B. J.; Bertino, J. R. *Mol. Pharmacol.* **1981**, *20*, 649.
(119) Dodgson, K. S.; Spencer, B.; Williams, K. *Biochem. J.* **1956**, *64*, 216.
(120) Bar, D. *Acta Pharm.* **1963**, *15*, 1.
(121) Hansch, C.; Muir, R. M.; Fujita,T.; Maloney, P. P.; Geiger, F.; Streich, M. *J. Am. Chem. Soc.* **1963**, *85*, 2817.
(122) Carotti, A.; Raguseo, C.; Campagna, F.; Langridge, R.; Klein, T. *Quant. Struct.-Act. Relat.* **1989**, *8*, 1.
(123) Tanaka, S.; Kato, T.; Takahashi, K.; Yamamoto, S. *Agric. Biol. Chem.* **1978**, *42*, 803.
(124) Hathaway, B. A.; Guo, Z. R.; Hansch, C.; Delcamp, T. J.; Susten, S. S.; Freisheim, J. H. *J. Med. Chem.* **1984**, *27*, 144.
(125) Kersters-Hilderson, H.; van Doorslaer, E.; De Bruyne, C. K. *Carbohydr. Res.* **1978**, *65*, 219.
(126) Schaeffer, H. J.; Johnson, R. N.; Odin, E.; Hansch, C. *J. Med. Chem.* **1970**, *13*, 452.
(127) Vickers, S.; Hebborn, P.; Moran, J. F.; Triggle, D. J. *J. Med. Chem.* **1969**, *12*, 491.
(128) Phillips, W. G.; Rejda-Heath, J. M. *Pestic. Sci.* **1993**, *38*, 1.
(129) Yu, L.; Wilkinson, C. F.; Anders, M. W. *Biochem. Pharmacol.* **1980**, *29*, 1113.
(130) Kraus, E.; Femfert, U. *Z. Physiol. Chem.* **1976**, *357*, 937.
(131) Schultz, T. W.; Cajina-Quezada, M. *Toxicol. Lett.* **1987**, 121.

(132) Fuller, R. W.; Marsh, M. M.; Mills, J. *J. Med. Chem.* **1968**, *11*, 397.
(133) Hansch, C.; Deutsch, E. W.; Smith, R. N. *J. Am. Chem. Soc.* **1965**, *87*, 2738.
(134) Seydel, J. K. *Mol. Pharmacol.* **1966**, *2*, 259.
(135) Warner, V. D.; Musto, J. D.; Sane, J. N.; Kim, K. H.; Grunewald, G. L. *J. Med. Chem.* **1977**, *20*, 92.
(136) Zimmerman, J. J.; Goyan, J. E. *J. Med. Chem.* **1971**, *14*, 1206.
(137) Laufer, S. A.; Augustin, J.; Dannhardt, G.; Kiefer, W. *J. Med. Chem.* **1994**, *37*, 1894.
(138) Hansch, C.; McClarin, J.; Klein, T.; Langridge, R. *Mol. Pharmacol.* **1985**, *27*, 493.
(139) van den Berg, G.; Bultsman, T.; Nauta, W. T. *Biochem. Pharmacol.* **1975**, *24*, 1115.
(140) Baker, B. R. *Design of Active-Site-Directed Irreversible Enzyme Inhibitors*; Wiley: New York, 1967; p 163.
(141) Caplow, M.; Jencks, W. F. *Biochemistry* **1962**, *1*, 883.
(142) Amshey, J. W.; Jindal, S. P.; Bender, M. L. *Arch. Biochem. Biophys.* **1975**, *169*, 1.
(143) Miyoshi, H.; Saitoh, I.; Iwamura, H. *Biochim. Biophys. Acta.* **1993**, *1143*, 23.
(144) Woodson, L. C.; Ames, M. M.; Selassie, C. D.; Hansch, C.; Weinshilboum, R. M. *Mol. Pharmacol.* **1983**, *24*, 471.
(145) Pocker, Y.; Beug, M. W. *Biochemsitry* **1972**, *11*, 698.
(146) Valchova, D.; L'Drobnica. *Collect. Czech. Chem. Commun.* **1966**, *31*, 997.
(147) Zannis, V. I.; Kirsch, J. F. *Biochemistry* **1978**, *17*, 2669.
(148) Derived from unpublished data from the National Cancer Institute.
(149) Cavallito, C. J.; Yun, H. S.; Kaplan, T.; Smith, J. C.; Foldes, F. F. *J. Med. Chem.* **1970**, *13*, 221.
(150) Suszkiw, J. B. *Biochem. Biophys. Res. Commun.* **1973**, *54*, 417.
(151) Morgenstern, L.; Recanatini, M.; Klein, T. E.; Steinmetz, W.; Yang, C. Z.; Langridge, R.; Hansch, C. *J. Biol. Chem.* **1987**, *22*, 10767.
(152) Williams, A.; Naylor, R. A.; Collyer, S. G. *J. Chem. Soc., Perkin Trans. 2* **1973**, 25.
(153) Alwani, D. W.; Noble, A. C.; Popp, F. D. *J. Med. Chem.* **1966**, *9*, 421.
(154) Campbell, N. R. C.; van Loon, J. A.; Sundaram, R. S.; Ames, M. M.; Hansch, C.; Weinshilboum, R. *Mol. Pharmacol.* **1987**, *32*, 813.
(155) Selassie, C. D.; Chow, M.; Hansch, C. *Chem.-Biol. Interact.* **1988**, *68*, 13.
(156) Seydel, J. K.; Kulkarni, V. M.; Coats, E. A.; Cordes, H. P. In *Pesticide Chemistry: Human Welfare and the Environment*; Miyamoto, Y., Ed.; Pergamon, Tarrytown, NY, 1983; Vol. 1, p 285.
(157) Stockdale, M.; Selwyn, M. J. *Eur. J. Biochem.* **1971**, *21*, 565.
(158) Cranfield, R.; Goodford, P. J.; Norrington, F. E.; Richards, W. H. G.; Sheppey, G. C.; Williams, S. G. *Br. J. Pharmacol.* **1974**, *52*, 87.
(159) Williams, A.; Woolford, G. *J. Chem. Soc., Perkin Trans. 2* **1972**, 272.
(160) Sanborn, J. R.; Fukuto, T. R. *J. Agric. Food Chem.* **1972**, *20*, 926.
(161) Cronenberger, L.; Dolfin, B.; Pacheco, H. *C. R. Acad. Sci. Paris, Ser. D* **1969**, *269*, 1334.
(162) Relimpio, A. M. *Gen. Pharmacol.* **1978**, *9*, 49.
(163) Darlington, W. A.; Partos, R. D.; Ratts, K. W. *Toxicol. Appl. Pharmacol.* **1971**, *18*, 542.
(164) Steurbaut, W.; Dejonckheere, W.; Kips, R. H. In *Crop Protection Agents*; McFarlane, N. R., Ed.; Academic Press: New York, 1977; p 79.
(165) Tollenaere, J. P.; Moereels, H.; van den Bossche, H. *Biochem. Parasites Host-Parasites Relat. Proc. Int. Symp.* **1976**, *2*, 629.
(166) Hansch, C.; Venger, B. J.; Panthananickal, A. *J. Med Chem.* **1980**, *23*, 459.
(167) Fukuto, T. R.; Metcalf, R. L. *J. Agric. Food Chem.* **1956**, *4*, 930.
(168) Metcalf, R. L.; Fukuto, T. R. *J. Econ. Entomol.* **1962**, *53*, 340.
(169) Metcalf, R. L.; Fukuto, T. R.; Frederickson, M. *J. Agric. Food Chem.* **1964**, *12*, 231.
(170) Fyfe, W. S. *J. Chem. Soc.* **1952**, 2018.
(171) White, G. A.; Thorn, G. D. *Pestic. Biochem. Physiol.* **1975**, *5*, 380.
(172) Hansch, C.; Klein, T.; McClarin, J.; Langridge, R.; Cornell, N. W. *J. Med. Chem.* **1986**, *29*, 615.
(173) Anderson, B. M.; Vasini, E. C. *Biochemistry* **1970**, *9*, 3348.
(174) Anderson, B. M.; Vercellotti, S. V.; Fisher, T. L. *Biochim. Biophys. Acta* **1974**, *350*, 135.
(175) Fonda, M. L.; Anderson, B. M. *J. Biol. Chem.* **1969**, *244*, 666.
(176) Mole, K. H.; Sheperd, D. M.; Watkins, W. J. *Arch. Inter. Pharmacodyn.* **1975**, *216*, 192.
(177) Ellul, B. M. J.; Moodie, R. B. *J. Chem. Soc. B* **1967**, 253.
(178) Zimmerman, J. J.; Goyan, J. E. *J. Med. Chem.* **1971**, *14*, 1206.
(179) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology;* American Chemical Society: Washington, DC, 1995; p 313.
(180) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology;* American Chemical Society: Washington, DC, 1995; p 462.
(181) Kwart, H.; Francis, P. S. *J. Am. Chem. Soc.* **1955**, *77*, 4907.
(182) Dyfverman, A.; Lindberg, B. *Acta Chem Scand.* **1950**, *4*, 878.
(183) Ferrier, R. J.; Overend, W. G.; Ryan, A. E. *J. Chem. Soc.* **1965**, 3484.
(184) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology;* American Chemical Society: Washington, DC, 1995; p 285.
(185) Muir, R. M.; Hansch, C. *Annu. Rev. Plant Physiol.* **1955**, *6*, 157.
(186) Robertson, R. E.; Stein, A.; Sugamori, S. E. *Can. J. Chem.*, **1966**, *44*, 685.
(187) Crumrine, D. S.; Shankweiler, J. M.; Hoffman, R. V. *J. Org. Chem.* **1986**, *51*, 5013.
(188) Burkhardt, G. N.; Horrex, C.; Jenkins, D. I *J. Chem. Soc.* **1936**, 1649.
(189) Hansch, C.; Zhang, L. *Drug Metab. Rev.* **1993**, *25*, 1.
(190) Tomila, E.; Hinshelwood, C. N. *J. Chem. Soc.* **1938**, 1801.
(191) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology;* American Chemical Society: Washington, DC, 1995; p 431.
(192) Hansch, C. In *Computer-Aided Molecular Design*: *Application in Agrochemical, Materials, and Pharmaceutics*; Holloway, M. K., Cox, H. H., Eds., Symposium Series 589; American Chemical Society: Washington, DC, 1995; p 281.
(193) Aldridge, W. N.; Davison, A. N. *Biochem. J.* **1952**, *51*, 62.
(194) van Hooidonk, C.; Ginjaar, L. *Rec. Trav. Chim.* **1967**, *86*, 449.
(195) Fukuto, T. R.; Metcalf, R. L. *J. Agric. Food Chem.* **1956**, *4*, 930.

CR9400976