PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Lead Optimization: Quantitative Structure-Activity Relations

A. Verloop

Phil. Trans. R. Soc. Lond. B 1981 **295**, 45-55
doi: 10.1098/rstb.1981.0124

Phil. Trans. R. Soc. Lond. B **295**, 45–55 (1981)
Printed in Great Britain

Lead optimization: quantitative structure–activity relations

By A. Verloop

Duphar B.V., P.O. Box 4, *Noordereinde* 63, *'s-Graveland, The Netherlands*

The optimization of bioactivity in a series is one of the aims in the design of crop protection chemicals. The approach of Hansch is the most developed model in this area. Here electronic, hydrophobic, and steric substituent constants are used for a quantitative analysis of the possible ways in which substitution may modulate bioactivity in a congeneric series (quantitative structure–activity relations; q.s.a.r.).

An essential prerequisite of this approach is the availability of reliable substituent constants. The present state of the art is reviewed, with emphasis on steric parameters where several new approaches have been developed recently, and based upon recent q.s.a.r. studies from the literature as well as from our own laboratories.

Q.s.a.r. may lead to significant regression equations, which can be used in different ways: the correlation between pesticidal activity and substituent properties provides the possibility of predicting the most favourable member of the class concerned. The resulting equations can also be given a physical interpretation in terms of the mode of action of the series studied. The scope and limitations of these applications are discussed.

Introduction

The optimization of bioactivity in a new series by variation of the substitution pattern is one of the aims in the design of crop protection chemicals. It was realized, early in the study of structure–activity relations guiding these optimization procedures, that more than one molecular property may influence bioactivity. This multiconditional character is caused by the fact that the biological activity of a crop protection chemical generally depends on several processes taking place in the biological object, including permeation and transport, (de)toxication and interaction with the receptor.

The change of hydrophobic properties may strongly modulate permeation and transport of the members of a series. These properties also determine largely the phenomenon of hydrophobic bonding to biological macromolecules, so that they may influence enzymic conversions and interactions with biological receptors. Hence hydrophobic constants may influence all the vital processes mentioned.

Differences in the electronic properties of substituents will influence the electron distribution in the parent molecule so that these properties will influence the energy of interaction between crop protection chemicals and (de)toxication enzymes or target receptors. In some cases electronic effects also influence permeation processes so that they, too, may be important for all the bioprocesses mentioned.

There is much evidence in the literature that the shape of molecules is of great importance for the interaction with their target receptors; the way in which substituents change this shape will therefore often influence this vital process appreciably. Examples of the importance as well as of the unimportance of steric properties for the rate of enzymic conversions have been described. But stereospecificity is mostly unimportant in (passive) permeation and transport processes. Hence steric properties will particularly influence the interaction of crop protection

TABLE 1. INFLUENCE OF SUBSTITUENT PARAMETERS ON BIOPROCESSES

bioprocess	parameters†		
	hydrophobic	electronic	steric
interaction with receptor	+	+	+
detoxication	+	+	−/+
permeation/transport	+	±	−

† +, Important; ±, less important; −, unimportant.

chemicals with the biological receptor and with stereospecific (de)toxication enzymes. The relative importance of the molecular properties for the different bioprocesses is summarized in table 1 (Verloop & Tipker 1976).

An approach with decisive importance for the study of structure–activity relations in a quantitative way (q.s.a.r.) was initiated by Hansch *et al.* (1962). In this method the way in which substituents influence the electronic, hydrophobic and steric properties of a parent molecule are defined in the following ways.

(i) For the electronic substituent effect the linear free energy parameter σ was taken, which is a measure of the electronic effect of a substituent s on a reaction centre of the parent molecule. This σ constant was developed, and defined in a reference system, in physical organic chemistry by Hammett (1940).

(ii) For the hydrophobic substituent effect Hansch defined the parameter π in an analogous way as a measure of the effect on the free energy of transfer of a substituent s from an aqueous phase to a lipophilic phase. Also, π was defined in a reference system, i.e. water–*n*-octanol.

(iii) For the steric substituent effect the linear free energy parameter E_s was used, which had been defined by Taft (1956), again in physical organic chemistry.

Hansch (1971) showed that the following correlation between biological activity of the members of a series of crop protection chemicals and these substituent parameters was generally to be expected:

$$\lg (1/c_s) = -a\pi^2 + b\pi + c\sigma + dE_s + e,$$

where c_s is the concentration of member s that gives a standard response in a standard time interval (e.g. e.d.$_{50}$, etc.), and a, b, c, d and e are constants. These constants can be obtained as regression coefficients by the statistical technique of multiple regression analysis. The occurrence of a π^2 term in the regression equation points to the possible presence of an optimum hydrophobicity in the series, probably caused by an optimum value for permeation through barriers (e.g. membranes) in the biological object studied. Squared terms are found occasionally also for the electronic and steric parameters, the latter suggesting the presence of an optimum fit to a biological receptor.

An essential prerequisite of this quantitative approach is the availability of reliable substituent constants. The σ constants and related electronic parameters are readily available from physical organic chemistry, and many of them have been applied in the design of crop protection chemicals (Verloop 1972; Hansch & Leo 1979). Considerable attention has also been given to hydrophobic parameters such as the π constants (Hansch & Leo 1979) and the related fragmentation (f) constants (Rekker 1977). The further development of the steric constants is, however, recent and for that reason, and also because our own group contributed to that development, the steric aspects of q.s.a.r. will be emphasized in this paper.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Steric aspects of q.s.a.r.

The E_s constants defined by Taft (1956) were the first steric substituent parameters used in quantitative structure–activity studies. Charton (1969) found that these constants are correlated with the *minimum width* of the Van der Waals radii of substituents. This discovery was further explored by Hansch, who calculated E_s constants for several substituents by using measured Van der Waals radii. Other authors have developed E_s constants corrected for

Table 2. Main steric substituent parameters in pesticide design

parameter	main character	authors
E_s	minimum width	Taft (1956)
E_s^c	minimum width	Hancock *et al.* (1961)
E_s^e	minimum width	Unger & Hansch (1976)
ν	minimum width	Charton (1973)
R	length	Bowden & Young (1970)
M_W	bulk	Hansch *et al.* (1973)
r_m	bulk	Hansch *et al.* (1973)
V_W	bulk	Moriguchi *et al.* (1976)
χ_v	bulk	Kier & Hall (1976)
$d_{s,\,min}$	deviation from ideal bulk	Simon *et al.* (1976)
B_1	minimum width	Verloop *et al.* (1976) (sterimol)
B_4	maximum width	Verloop *et al.* (1976) (sterimol)
L	length	Verloop *et al.* (1976) (sterimol)

electronic influences, such as the E_s^c, E_s^e, and ν parameters, all of which essentially measure the minimum widths of substituents (table 2) (Verloop & Tipker 1977). Bowden & Young (1970) have used the constant R for the *length* of a substituent; the R constants were measured from molecular models. Several parameters for *general steric bulk* have been introduced in q.s.a.r., i.e. molecular mass (M_w), molar refractivity (r_m), Van der Waals volume V_W, molecular connectivity index (χ_v) and minimal steric difference ($d_{s,\,min}$) (table 2) (Verloop & Tipker 1977).

The common feature of all of these new steric parameters is that only one aspect of the shape of substituents is described, i.e. minimum width, length or general steric bulk. This must be considered a questionable approach for non-spherical substituents. A consideration of molecular models of substituents used in the design of crop protection chemicals readily leads to the conclusion that in many cases the width of a substituent may vary considerably in different directions. Also, the relative variations in length and widths of substituents can deviate considerably.

In our group we therefore developed, by means of the computer program sterimol, a number of steric parameters representing the widths and the lengths of substituents in several directions, measured by using C.P.K. atomic models. For flexible substituents the minimum energy conformations were adopted (Verloop *et al.* 1976). In this paper only the three most important sterimol parameters will be discussed, namely the length, L, of a substituent, measured in the direction in which the group is attached to the parent molecule; the minimum width, B_1, and the (approximate) maximum width, B_4, measured in directions perpendicular to the L axis (table 2) (Verloop & Tipker 1977).

Applications of q.s.a.r.

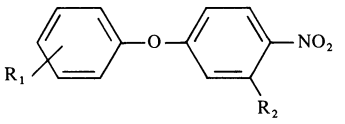
A recent study on the inhibition of the Hill reaction in chloroplasts by the herbicidal 4-nitrophenylethers can serve as a first example (Van den Berg & Tipker 1981). The results of the q.s.a.r. analysis are given in table 3.

In the q.s.a.r. multi-regression analyses discussed in this paper, n is the number of compounds in the series, r is the correlation coefficient, s is the standard deviation and F is the value from the F-test which, after tabular interpretation, indicates the significance of the correlation found. The value r^2 is interpreted as the fraction of the variance in the biological data that is attributable to the regression equation. In the tables the regression equations are given in an abbreviated form. The complete equation in the first column of table 3, for example, is

$$pI_{50} = 0.512\pi - 0.071\pi^2 + 4.340.$$

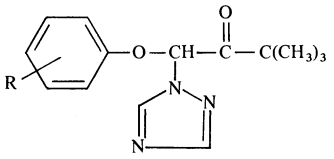
In the tables, only the best regression equation obtained is given; for more details, e.g. about the individual members of the series, the original references should be consulted. From the results

TABLE 3. INHIBITION OF HILL REACTION IN CHLOROPLASTS FROM SPINACH LEAVES BY 4-NITROPHENYLETHERS (pI_{50})



$R_1 = n$; $R_2 = H$	R_1 = 2,4-diCl; $R_2 = n$	$R_1, R_2 = n$
$+0.512\pi$	.	$+0.507\pi$ (R_1)
$-0.071\pi^2$	.	$-0.070\pi^2$ (R_2)
.	$-0.239B_4$	$-0.259B_4$ (R_2)
+4.340	+5.107	+4.594
$n = 18$	$n = 7$	$n = 24$
$r = 0.965$	$r = 0.928$	$r = 0.961$
$s = 0.159$	$s = 0.084$	$s = 0.140$
$F = 107.64$	$F = 63.98$	$F = 92.01$

TABLE 4. FUNGICIDAL ACTIVITY OF PHENOXY-TRIAZOLYLMETHANES TO *ERYSIPHE CICHORACAERUM* (lg pI_{50})



$-0.35B_4$	$-0.35B_4$
$+0.55\pi$	$+1.80R_M$
+5.38	+5.46
$n = 24$	$n = 26$
$r = 0.868$	$r = 0.847$
$s = 0.373$	$s = 0.389$
$F = 31.9$	$F = 28.0$

given in table 3 it is clear that bioactivity is completely governed by the hydrophobic π constants if the structural variation is restricted to the substituents in the phenyl ring at the left side; here the π^2 term indicates an optimum hydrophobicity, probably caused by an optimum permeation into the chloroplasts. But substitution in the other ring, *ortho* to the NO_2 group, is only dependent on the maximum width of the substituents. Both subseries can be combined in one regression equation. This example illustrates the necessity of analysing subseries separately to obtain the best understanding of the q.s.a.r.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Another study is illustrated in table 4. Here the best regression equations are given, which were obtained by Krämer *et al.* (1979) in q.s.a.r. studies on the fungicidal activity of substituted phenoxy triazolylmethanes to *Erysiphe cichoracaerum*. The compound with R = *p*-Cl is the new Bayer fungicide triadimefon (trade name Bayleton). No significant correlations were found when applying electronic parameters or the steric E_s constants, but table 4 shows that significant results

TABLE 5. COMPARISON OF CORRELATION COEFFICIENTS AND OF QUALITY OF DIFFERENT SETS IN THE FOUR STUDIES

	ligands–papain $n = 13$	carbamates $n = 29$	phenylbenzoylureas $n = 26$	haptens–antibody $n = 29$	haptens–antibody $n = 36$
parameters					
r_m	0.935	0.889	0.659	0.893	0.714
χ	0.938	0.899	0.701	0.842	0.723
ν	0.878	0.933	0.638	0.960	—
$d_{s,\ min}$	0.924	0.933	0.697	0.944	0.655
STERIMOL	0.918	0.940	0.848	0.945	0.953
range of STERIMOL parameters					
L/B_1	2.0–2.9	2.0–3.0	2.0–9.0	2.0–3.0	2.0–5.5
B_4/B_1	1.0–2.1	1.0–2.3	1.0–7.0	1.0–2.1	1.0–4.0

were obtained with the STERIMOL parameter B_4, combined with either the hydrophobic π constant or an experimental hydrophobic R_M parameter.

To obtain more insight into applicability of the different steric constants mentioned in the introduction, we carried out comparative studies (Verloop & Tipker 1976). Table 5 summarizes the results of the following q.s.a.r. studies.

1. The binding to the enzyme papain of a series of ligand molecules with the general structure $RC_6H_4OCOCH_2NHSO_2CH_3$, and R at the *m*- and *p*-positions.

2. The inhibition of acetylcholinesterase by substituted phenyl-*N*-methylcarbamates.

3. The inhibition of the moulting of larvae of *Pieris brassicae* L. by *p*-substituted phenyl-2,6-dichlorobenzoylureas. These compounds are analogues of the new Duphar chitin synthesis-inhibiting insecticide diflubenzuron (trade name Dimilin).

4. The hapten–antibody interaction of substituted benzoic acids with X-ovalbumin antigen.

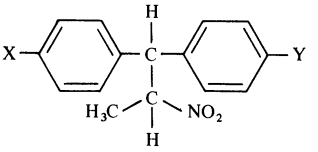
In table 5 the correlation coefficients are given of the most significant regression equations when different steric parameters are used, in most cases together with hydrophobic and electronic substituent constants. These coefficients are compared with the ranges of values for L/B_1 and B_4/B_1. These ratios are measures of the deviation of the substituents from the spherical shape in length and width. It can be concluded that in the studies on the ligands–papain interactions, on the effect of the carbamates on acetylcholinesterase and on the hapten–antibody interaction with a limited set of 29 derivatives, there is not much difference between the correlation coefficients when the bulk parameters r_m, χ_v and $d_{s,\ min}$, the minimum width parameter ν, or the STERIMOL constants are applied. In all these cases the maximum deviation of the substituents from the spherical shape was very limited. However, in the two other studies on the insecticidal activity of the phenylbenzoylureas and the hapten–antibody interaction with the extended set of 36 derivatives, there was much more variation in the shape of the substituents and here it was found that the STERIMOL parameters L and B_4 gave by far the highest correlation coefficients.

These results are supported by studies published recently. Lee *et al.* (1977) carried out a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

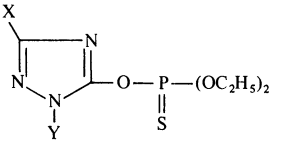
q.s.a.r. study on the DDT-related insecticide Prolan and its derivatives; they analysed the role of several types of bulk parameters and of the STERIMOL constants in the regression equations of the insecticidal activity of 25–35 members on five different insect species. The correlation coefficients of the best equations obtained in each case are assembled in table 6. It is clear that in most cases improved equations were obtained by using the STERIMOL parameters L and B_4, together with electronic and hydrophobic constants. For the black blowfly this was the only

TABLE 6. EFFECT OF STERIC PARAMETERS ON INSECTICIDAL ACTIVITY OF PROLAN ANALOGUES (l.c.$_{50}$ or l.d.$_{50}$)



insect	correlation coefficients of best equation with: 'bulk' parameters	STERIMOL parameters
housefly (S_{NAIDM})	0.893	0.897
housefly (R_{SP})	0.788	0.861
Culex larvae	0.696	0.877
Anopheles larvae	0.676	0.923
black blowfly	n.s.	0.856

TABLE 7. INSECTICIDAL ACTIVITY TO COLORADO POTATO BEETLES OF *O,O*-DIETHYL *O*-TRIAZOLYLPHOSPHOROTHIOATES ($-$lg l.d.$_{50}$)



$+0.609B_4^X$	$+0.065r_m^X$
$-1.224B_1^Y$	.
$-0.255B_4^Y$	.
-1.244	-2.775
$n = 21$	$n = 21$
$r = 0.965$	$r = 0.722$
$s = 0.173$	$s = 0.430$
$F = 76.14$	$F = 20.75$

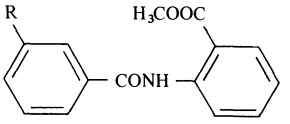
procedure, of those tried, giving a significant result. In this study there was again a large variation in the geometry of the substituents, i.e. $L/B_1 = 2.0$–8.5, and $B_4/B_1 = 1.0$–5.7.

Gozzo *et al.* (1980) studied the insecticidal activity of triazolyl diethoxythiophosphates to Colorado potato beetles (table 7). The first column shows that with the substituent constant B_4 for substituents X, and constants B_1 and B_4 for substituents Y, a very significant regression equation was obtained. With the bulk parameter r_m for substituents X a significant correlation was also obtained although it was considerably less exact. The possible influence of hydrophobic and electronic substituent constants was also tested, but no correlation was detected. In this study L/B_1 was 1.7–4.0, and B_4/B_1 was 1.0–3.0.

A last example is the q.s.a.r. analysis of the fungicidal activity of *N*-benzoylanthranilates against powdery mildew of cucumber caused by *Sphaerotheca fuliginea*. Kirino *et al.* (1980*a*) examined various combinations of substituent parameters. No significant correlation was found

when applying the hydrophobic, π, and electronic, σ, constants. The use of the steric parameters E_s, r_m, V_u (another bulk parameter) singly or together with π and σ constants, also did not result in significant regression equations. But good results were obtained with the STERIMOL parameter L and in particular with B_4 (table 8). From the L^2 term it was calculated that in this series the optimum length of the substituents was 0.474 nm, possibly because of the demands for an optimum fit to the target site. Here again there was a relatively large deviation of the substituents from the spherical shape: the maximum values for L/B_1 and B_4/B_1 were 4.5 and 3.2 respectively.

TABLE 8. FUNGICIDAL ACTIVITY OF *N*-BENZOYLANTHRANILATES TO *SPAEROTHECA FULIGINEA* (pI_{50})



.	$-0.114L^2$
.	$+1.082L$
$-0.568B_4$	$-0.715B_4$
$+5.371$	$+3.349$
$n = 14$	$n = 14$
$r = 0.885$	$r = 0.942$
$s = 0.289$	$s = 0.229$
$F = 43.30$	$F = 26.02$

From the studies discussed it can be concluded that in those examples of q.s.a.r. where steric requirements play a role, the STERIMOL parameters can best accommodate great variations in steric substituent properties. This is especially so whenever the maximum width parameter B_4 or the length L of the substituents are important. When the minimum width is significant the STERIMOL B_1 constant as well as parameters like E_s or the Charton constant ν are useful, the choice between these constants depending on the specific series studied.

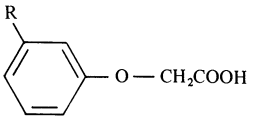
SCOPE AND LIMITATIONS OF Q.S.A.R.

The use of multiple regression equations in q.s.a.r. can be divided into two broad categories, each with its own scope and limitations.

In the first place, these relations may be considered as pure correlations between bioactivity and substituent constants. The correlations can be used for several purposes, which are all of great importance for the discovery of new crop protection chemicals by industry. The most important aspect is the possibility of predicting the activity of members of a series that are not yet synthesized. This predictive power is dependent on the quality of the regression equations obtained. As an example I shall take our recent study on plant growth regulators. The analysis of the plant cell elongation by substituted phenoxyacetic acids (Hansch *et al.* 1963) was the very first q.s.a.r. analysis according to the Hansch approach. In this study, 25 *ortho*, *meta* and *para* substituted members were used, of which three were not included in the regression analysis, i.e. *m*-OH, *m*-SO_2CF_3 and *m*-COOH. In a new study we restricted ourselves to the 20 *meta* substituted derivatives and eliminated only the *m*-COOH derivative because of its uncertain degree of ionization, so that 19 compounds were analysed (table 9).

Our version of the original Hansch analysis, with π, π^2 and σ substituent parameters, is

TABLE 9. PLANT GROWTH REGULATION ACTIVITY OF *m*-SUBSTITUTED PHENOXYACETIC ACIDS IN THE CELL ELONGATION TEST ($-\lg C$)



$+0.89\pi$	$+0.93\pi$	$+1.12\pi$	$+0.91\pi$	$+1.04\pi$	$+0.93\pi$
.	$+1.15\sigma^m$	$+1.32\sigma^m$	$+0.83\sigma^m$	$+0.59\sigma^m$	$+1.05\sigma^m$
.	.	$-0.29\pi^2$	$-0.02r_m^2$	$-0.67L^2$	.
.	.	.	$+0.29r_m$	$+4.78L$	$-0.54d_{s,\,min}$
$+4.46$	$+4.07$	$+4.15$	$+3.36$	-3.87	$+4.97$
$n = 19$	$n = 19$	$n = 19$	$n = 19$	$n = 19$	$n = 19$
$r = 0.738$	$r = 0.808$	$r = 0.832$	$r = 0.873$	$r = 0.935$	$r = 0.930$
$s = 0.648$	$s = 0.584$	$s = 0.568$	$s = 0.516$	$s = 0.376$	$s = 0.376$
$F = 20.36$	$F = 15.03$	$F = 11.20$	$F = 11.21$	$F = 24.19$	$F = 32.00$

TABLE 10. PREDICTIVE POWER OF REGRESSION EQUATIONS FOR AUXIN ACTIVITY OF PHENOXYACETIC ACIDS

		$-\lg C$ (calc.)			
substituent	$-\lg C$ (obs.)	π^2 – eq.	r_m – eq.	L – eq.	$d_{s,\,min}$ – eq.
$NHCOCH_3$	< 3	3.6	3.1	**2.2**	**2.4**
n-C_4H_9	< 3	5.1	3.9	**1.8**	4.8
$NHCOC_6H_5$	< 3	4.8	**−7.2**	**−7.5**	**1.1**
C_6H_5	< 3	5.3	**1.1**	**1.3**	4.4
OCF_3	< 3	5.6	5.5	5.3	4.3

given in the first three columns of this table. The highest correlation coefficient, $r = 0.832$, was obtained when all three terms were used; this is in accordance with the analysis by Hansch *et al.* (1963). In the next column the improvement is shown when the steric bulk parameter r_m is added. With the minimum width parameters E_s and B_1 no improvement of the original equation was obtained. But the application of the STERIMOL length parameter L led to a considerably better correlation coefficient of 0.935. A similar improvement was obtained with the $d_{s,\,min}$ parameter, as shown in the last column. In both cases almost 90% of the variance in the data was explained, whereas in the original Hansch equation only about 70% was explained. Hansch and coworkers did not analyse the possible importance of steric influences because, from previous studies on modes of action, these were considered improbable. Nevertheless, in our new study these influences were found to be very significant.

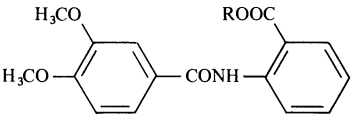
In table 10 we have summarized the results of testing the predictive power of the different regression equations by comparing the observed bioactivities with the calculated activities for five *meta* substituted derivatives that had not been used for the determinations of the equations. The calculated values in bold type are in agreement with the observed data, so that the correct predictions were as follows: for the π^2 equation, none; for the r_m and $d_{s,\,min}$ equations, two out of five; and for the L equation, four out of five. The OCF_3-substituted derivative was the only one that was predicted wrongly in all cases. Other examples where the use of STERIMOL parameters led to successful predictions are the q.s.a.r. studies by Gozzo *et al.* (1980), summarized in table 7, and those by Iwamura *et al.* (1980).

Another important use of q.s.a.r. for industrial research is its help in planning of programmes

of synthesis: for the proper application it is important that the substituent parameters are varied in such a way that the 'spanned substituent space' is as great as possible. This is also vital for the predictive power of the equations obtained, which is generally much better for interpolated compounds than for extrapolated ones.

Q.s.a.r. can also contribute to the understanding of selectivity and thus to the development of

TABLE 11. FUNGICIDAL ACTIVITY OF *N*-BENZOYLANTHRANILIC ESTERS TO *ERYSIPHE GRAMINIS* (pI_{50})



$-5.624R_M^2$	$-5.27R_M^2$	$-4.58R_M^2$	$-5.12R_M^2$
$-0.902\sigma^*$	$-0.88\sigma^*$	$-0.72\sigma^*$	$-1.47\sigma^*$
$-1.287X$	$-1.53X$	$-3.29B_1$	$-3.37B_1$
$+5.410$	$+5.39$	$+10.33$	$+0.40B_4$
			$-0.32L$
			$+10.54$
$n = 21$	$n = 23$	$n = 23$	$n = 23$
$r = 0.963$	$r = 0.913$	$r = 0.887$	$r = 0.965$
$s = 0.207$	$s = 0.383$	$s = 0.434$	$s = 0.235$
$F = 72.45$	$F = 31.74$	$F = 23.29$	$F = 43.90$

selective crop protection chemicals, for example by the comparison of regression equations obtained for the desired bioactivity and for side effects. Examples for the development of chemicals with low mammalian toxicity are published by Verloop & Tipker (1976), and Gozzo *et al.* (1980). The use of q.s.a.r. in the study of fungicidal selectivity has been discussed by Verloop & Tipker (1976).

Other examples illustrating the scope of q.s.a.r. for industrial research are given by Verloop (1972); it should, however, be realized that the Hansch approach has serious limitations too: in several cases no very significant regression equations are obtained with the available substituent constants. Attempts are made to improve this situation, e.g. by omitting certain members of the series, or by using so-called dummy or indicator parameters to cope with the deviating behaviour of members with common characteristics. An example is a recent study of Kirino *et al.* (1980*b*) on the fungicidal activity of *N*-benzoylanthranilic esters to powdery mildew of barley (table 11). In the first column the best results obtained by Kirino *et al.* are summarized. The hydrophobic experimental constant R_M and the electronic σ^* parameter for aliphatic substituents influence the correlation, together with the indicator parameter X. This variable X is taken as unity for alkyl esters with α-branching, and zero for the other esters. But the authors omitted the cyclopentyl and cyclohexyl derivatives. The inclusion of these members considerably diminishes the significance of the regression equation, as shown in the second column. The indicator parameter is determined by steric factors, as shown in columns three and four. The STERIMOL parameter B_1 has the most important influence, probably related to the α-branching effect, but the addition of the L and B_4 parameters leads to the best regression equation, which is even slightly better for all 23 members than the equation of Kirino *et al.* for 21 members.

From this example and from the studies discussed earlier it may be concluded that the improvement of the present sets of substituent parameters, e.g. the development of the STERIMOL

approach, can strengthen further the quality of q.s.a.r. This is especially so in the area of steric influences on the q.s.a.r. of more specific crop protection chemicals, an important aspect of future industrial research. In the STERIMOL approach, the molecular geometry measurements are based upon the assumption that minimum energy conformations, i.e. extended forms, are present. However, this is problematic for flexible substituents so that here also there is scope for improvement.

The second broad category of uses is based upon the character of the Hansch approach that the substituent parameters have a physical meaning. It is therefore possible and even tempting to give the regression equations a physical interpretation in terms of the mode of action of the relevant series. This implies the sometimes dangerous step from correlation towards causality. This use, which is rather popular in several non-industrial laboratories, has been critically reviewed elsewhere (Verloop 1972). The main conclusion was that a regression equation can be interpreted in several ways so that further experiments, e.g. toxicodynamic approaches (Welling 1977), with the use of independent methods are necessary to choose between the different interpretations. Multiple regression equations can at most provide a starting point in terms of a working model.

In conclusion, the most important applications of q.s.a.r. in its present state of development are its help in planning synthetic optimization programmes of new classes of crop protection chemicals, in interpreting the results obtained, and in possibly predicting more active members of the series studied. In this way q.s.a.r. may influence the present success rate, now estimated to be of the order of 1:12000. Approximately two-thirds of these 12000 compounds needed to develop one useful new crop protection chemical are synthesized in lead optimization programmes. It is obvious that the successes of q.s.a.r. approaches so far, although modest, justify their contribution to the research efforts of the modern crop protection chemical industry. It is therefore not surprising that in spite of initial scepticism q.s.a.r. studies are now an integral part of these efforts.

I am indebted to Mr J. Tipker for valuable discussions and for new q.s.a.r. calculations.

REFERENCES (Verloop)

Bowden, K. & Young, R. C. 1970 Structure–activity relations. I. A series of antagonists of acetylcholine and histamine at the postganglionic receptors. *J. med. Chem.* **13**, 225–230.

Charton, M. 1969 The nature of the *ortho* effect. II. Composition of the Taft steric parameters. *J. Am. chem. Soc.* **91**, 615–618.

Charton, M. 1973 Substituent effects in nonaromatic unsaturated systems. *Prog. phys. org. Chem.* **10**, 81–104.

Gozzo, F., Masoero, M., Quadrelli, L. & Zagni, A. 1980 A novel series of *O,O*-diethyl *O*-(1,3-disubstituted-1,2,4-triazolyl-*S*-yl) phosphorothioates. II. Quantitative structure–activity relationships. *Pestic. Sci.* **11**, 314–323.

Hammett, L. P. 1940 *Physical organic chemistry*. New York: McGraw-Hill.

Hancock, C. K., Meyers, E. A. & Yager, B. J. 1961 Quantitative separation of hyperconjugation effects from steric substituent constants. *J. Am. chem. Soc.* **83**, 4211–4213.

Hansch, C. 1971 Quantitative structure–activity relationships in drug design. In *Drug design* (ed. E. J. Ariëns), vol. 1, pp. 271–342. New York: Academic Press.

Hansch, C. & Leo, A. 1979 *Substituent constants for correlation analysis in chemistry and biology*. New York: John Wiley & Sons.

Hansch, C., Leo, A., Unger, S. H., Kim, K. H., Nikaitani, D. & Lien, E. 1973 'Aromatic' substituent constants for structure–activity correlations. *J. med. Chem.* **16**, 1207–1216.

Hansch, C., Maloney, P. P., Fujita, T. & Muir, R. M. 1962 Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature, Lond.* **194**, 178–180.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, F. & Streich, M. 1963 The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. *J. Am. chem. Soc.* **85**, 2817–2824.

Iwamura, H., Fujita, T., Koyama, S., Koshimizu, K. & Kumazowa, Z. 1980 Quantitative structure–activity relationship of cytokinin-active adenine and urea derivatives. *Phytochemistry* **19**, 1309–1319.

Kier, L. B. & Hall, L. H. 1976 *Molecular connectivity in chemistry and drug research*. New York: Academic Press.

Kirino, O., Yamamoto, S. & Kato, T. 1980*a* Fungicidal activity of *N*-benzoylanthranilates and related compounds. *Agric. biol. Chem.* **44**, 2143–2147.

Kirino, O., Yamamoto, S. & Kato, T. 1980*b* Structure–activity relationships of fungicidal *N*-benzoylanthranilic esters. *Agric. biol. Chem.* **44**, 2149–2153.

Krämer, W., Büchel, K. H., Meiser, W., Brandes, W., Kaspers, H. & Scheinpflug, H. 1979 Triazolyl-*O*,*N*-acetals, chemistry, activity and structure. In *Advances in pesticide science* (ed. H. Geissbühler), vol. 2, pp. 274–279. Oxford: Pergamon Press.

Lee, A.-H., Metcalf, R. L., Williams, J. W., Hirwe, A. S., Sanborn, J. R., Coats, J. R. & Fukuto, T. R. 1977 Structure–activity correlations of the insecticide Prolan and its analogs. *Pestic. Biochem. Physiol.* **7**, 426–436.

Moriguchi, I., Kanada, Y. & Komatsu, K. 1976 Van der Waals volume and the related parameters for hydrophobicity in structure–activity studies. *Chem. pharm. Bull., Tokyo* **24**, 1799–1806.

Rekker, R. F. 1977 *The hydrophobic fragmental constant*. Amsterdam: Elsevier.

Simon, Z., Chiriac, A., Motoc, I., Holban, S., Ciubotaru, D. & Szabadai, Z. 1976 Receptor site mapping. Search strategy of standard for correlation with minimal steric differences. *Stud. Biophys.* **55**, 217—226.

Taft, R. W., Jr 1956 Separation of polar, steric, and resonance effects in reactivity. In *Steric effects in organic chemistry* (ed. M. S. Newman), pp. 556–675. New York: John Wiley & Sons.

Unger, S. H. & Hansch, C. 1976 Quantitative models of steric effects. *Prog. phys. org. Chem.* **12**, 91–118.

Van den Berg, G. & Tipker, J. 1981 Quantitative structure–activity relationships of the inhibition of photosynthetic electron flow by substituted diphenylethers. *Pestic. Sci.* (Submitted.)

Verloop, A. 1972 The use of linear free energy parameters and other experimental constants in structure–activity studies. In *Drug design* (ed. E. J. Ariëns), vol. 3, pp. 133–187. New York: Academic Press.

Verloop, A., Hoogenstraaten, W. & Tipker, J. 1976 Development and application of new steric substituent parameters in drug design. In *Drug design* (ed. E. J. Ariëns), vol. 7, pp. 165–207. New York: Academic Press.

Verloop, A. & Tipker, J. 1976 Use of linear free energy related and other parameters in the study of fungicidal selectivity. *Pestic. Sci.* **7**, 379–390.

Verloop, A. & Tipker, J. 1977 A comparative study of new parameters in drug design. In *Biological activity and chemical structure* (ed. J. A. Keverling Buisman), pp. 63–81. Amsterdam: Elsevier.

Welling, W. 1977 Dynamic aspects of insect–insecticide interactions. *A. Rev. Ent.* **22**, 53–78.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES