

## Quantitative Structure-Activity Relationships. 2. A Mixed Approach, Based on Hansch and Free-Wilson Analysis

Hugo Kubinyi

Research Institute of Knoll AG, Chemische Fabriken, D-6700 Ludwigshafen/Rhein, Germany. Received June 27, 1975

Based on the theoretical and numerical equivalence of Hansch's linear multiple regression model and the modified Free-Wilson model a mixed approach is developed. The mixed approach is a combination of both models which makes use of the advantages of each model and widens the applicability of Hansch and Free-Wilson analysis. The Free-Wilson approach now is applicable also in the case of parabolic dependence of biological activity on a particular physical property, e.g.,  $\log P$  or  $\pi$ . A rational explanation is given for the use of dummy variables in Hansch equations and the derivation of Hansch correlations for de novo group contributions obtained from Free-Wilson analysis. Some examples illustrate the mixed approach and demonstrate its usefulness to establish biologically meaningful structure-activity relationships.

The nonparabolic form of Hansch's linear multiple regression model<sup>1-3</sup> (e.g., eq 1 or, more generally expressed,

$$\log 1/C = k_1\pi + k_2\sigma + k_3E_s + k_4 \quad (1)$$

eq 2) and a modified form<sup>4,5</sup> of Free-Wilson's additive

$$\log 1/C = \sum_j k_j\phi_j + k \quad (2)$$

$\phi_j$  = structural parameters like  $\pi$ ,  $\sigma$ , or  $E_s$ .

model<sup>6</sup> (eq 3)

$$\log 1/C = \sum_i a_i + \mu \quad (3)$$

where  $a_i$  = group contribution of substituent  $X_i$ , based on  $a_H = 0.00$  and  $\mu = \log 1/C$  calcd of the unsubstituted compound<sup>5</sup> have been shown to be theoretically interrelated<sup>7,8</sup> and numerically equivalent<sup>8</sup> (a detailed description of this modified Free-Wilson model is given by Fujita and Ban;<sup>5</sup> note that Cammarata and Yau<sup>4</sup> defined  $\mu = \log 1/C$  obsd of the unsubstituted compound, which gives different results; compare ref 8).

This equivalence is due to the fact that the individual group contributions can be interpreted as a weighted sum of several physical properties  $\phi_j$  of the substituents (eq 4).

$$a_i = \sum_j b_j\phi_j \quad (4)$$

Substitution of eq 4 into the Free-Wilson type eq 3 gives, after appropriate transformation,<sup>8</sup> eq 5, which is a Hansch type equation.

$$\log 1/C = b_1\sum\phi_1 + b_2\sum\phi_2 + \dots + b_n\sum\phi_n + \mu \quad (5)$$

Biological activity, expressed as  $\log 1/C$ , often is parabolically dependent on lipophilic character, expressed as  $\log P$  or  $\pi$ .<sup>1-3,9</sup> In these cases the linear multiple regression model takes the form of eq 6. The additive Free-Wilson

$$\log 1/C = k_1\pi^2 + k_2\pi + k_3\sigma + k_4E_s + k_5 \quad (6)$$

model is not appropriate in the case of significant parabolic dependence of  $\log 1/C$  values on the lipophilic character of substances<sup>8,10</sup> (note that Singer and Purcell<sup>10</sup> demonstrated that the interaction model of Bocek and Kopecky<sup>11,12</sup> corresponds to the parabolic form of Hansch's linear multiple regression model, eq 6).

Based on the equivalence of eq 1 and 3 the Hansch eq 1, 2, and 6 can be transformed in three different ways.

Firstly, the nonparabolic part of eq 6 can be substituted by  $\sum a_i + \mu$  (eq 3) to give eq 7. Equation 7 can be in-

$$\log 1/C = k_1\pi^2 + \sum_i a_i + \mu \quad (7)$$

terpreted as a Free-Wilson equation with an additional term  $k_1\pi^2$  to account for parabolic dependence of  $\log 1/C$  on lipophilic character or as a Bocek-Kopecky<sup>11,12</sup> like equation,  $k_1\pi^2$  being the interaction term, or as a "mixed approach" with a Hansch part  $k_1\pi^2$  and a Free-Wilson part  $\sum a_i$ .

Secondly, a linear combination of a Free-Wilson part  $\sum a_i$  for substituents  $X_i$  (eq 3) and a Hansch part  $\sum k_j\phi_j$  for substituents  $Y_i$  (eq 2) gives a "mixed equation", eq 8.

$$\log 1/C = \sum_i a_i + \sum_j k_j\phi_j + k' \quad (8)$$

Thirdly, eq 3 and 6 (or eq 7 and 8) can be combined to eq 9 which is a mixed approach with a Free-Wilson part

$$\log 1/C = k_1\pi^2 + \sum_i a_i + \sum_j k_j\phi_j + k' \quad (9)$$

$\sum a_i$  for substituents  $X_i$ , a Hansch part  $\sum k_j\phi_j$  for substituents  $Y_i$ , and a term  $k_1\pi^2$  accounting for parabolic dependence of  $\log 1/C$  values on lipophilic character (note that  $\pi$  in  $k_1\pi^2$  must be  $\pi_x + \pi_y$ ).

Equations 7, 8, and 9 look rather complex at first sight, but in reality they are simple variations of Free-Wilson's additive model which make this approach more generally applicable; the Hansch part in eq 8 and 9 gives a reduction of the number of variables needed, if for one definite group of substituents a Hansch equation can be derived. The term  $k_1\pi^2$  in eq 7 and 9 makes the Free-Wilson approach now applicable to all problems where nonadditivity of group contributions is caused by parabolic dependence of  $\log 1/C$  on  $\log P$  or  $\pi$ . Since values of  $\log P$ <sup>13</sup> and  $\pi$ <sup>14,15</sup> are known for a great number of compounds and substituents and can be estimated<sup>3,13</sup> for new compounds and substituents, the use of  $k_1\pi^2$  in eq 7 and 9 gives no restrictions for the general applicability of the mixed approach (some attention should be paid to possible non-additivities of  $\pi$  values;<sup>13,16,17</sup> in cases of greater deviations from additivity experimental  $\log P$  values should be used).

Craig and Hansch<sup>18,19</sup> discussed the different requirements for application of the linear multiple regression model and the additive model. In most cases the Hansch approach is the more general and useful model but there are also limitations for this approach; for certain groups of compounds today only the Free-Wilson approach can give correlations between chemical structure and biological activity (for examples see ref 5 and 20). A further limitation of the Hansch approach comes from little structural variation in a definite position of the molecule; no meaningful Hansch correlations can be derived if only two or three substituents are in a position where they contribute to biological activity in a different manner related to substituents in other positions of the molecule. In these instances it might be useful to combine both models in the

Table I. Hansch Correlations for de Novo Group Contributions

Cammarata and Yau, <sup>4</sup> tetracyclines, ring D substituents		
$a_i = 0.87 \sigma^2 - 0.57 r_v + 0.84$	$n = 7; r = 0.93$	
Fujita and Ban, <sup>5</sup> substrates of phenylethanolamine N-methyltransferase, aromatic substituents		
$a_i = 0.743 \pi + 0.881 \sigma_1 + 0.323$	$n = 5; r = 0.906$	
$a_i = 0.903 \pi - 1.030 \sigma_2 + 0.214$	$n = 5; r = 0.954$	
Craig, <sup>21</sup> 2-phenylquinoline-4-carbinols, substituents		
$R_4: a_i = 0.220 \pi + 0.626 \sigma_m - 0.232$	$n = 7; r = 0.895$	
$R_7: a_i = 1.811 \sigma_p - 0.010$	$n = 5; r = 0.981$	
$a_i = 0.998 \pi - 0.198$	$n = 5; r = 0.942$	
$R_8: a_i = 0.959 \pi - 0.395$	$n = 4; r = 0.966$	
Craig and Hansch, <sup>19</sup> phenanthreneaminoalkylcarbinols, substituents		
$R_3: R_3 \text{ vs. } \pi$	$n = 6; r = 0.971$	
$R_3 \text{ vs. } \sigma$	$n = 6; r = 0.910$	
$R_6: R_6 \text{ vs. } \pi$	$n = 6; r = 0.910$	
$R_6 \text{ vs. } \sigma$	$n = 6; r = 0.885$	

form of eq 8 or 9, e.g., if one considers a group of compounds with aromatic substituents and substituents in a side chain; the aromatic substituents may be correlated with the Hansch parts of eq 8 or 9 and the substituents of the side chain may be correlated with the Free-Wilson parts of eq 8 or 9, if no significant correlations with structural parameters can be obtained for these substituents.

It is surprising that, to my knowledge, no use has been made of eq 7 until today.

On an empirical basis two different applications of eq 8 and 9 have been made in the past; these applications are the derivation of Hansch equations for de novo group contributions and the use of certain dummy variables (indicator variables) in Hansch analysis.

Cammarata and Yau<sup>4</sup> and Fujita and Ban<sup>5</sup> were the first ones to demonstrate that analysis of Free-Wilson group contributions by the linear multiple regression model may lead to Hansch correlations which allow predictions of log 1/C values for compounds with substituents outside the original Free-Wilson matrix. The Hansch equations derived by these groups and some other successful correlations of Free-Wilson group contributions with structural parameters are given in Table I.

A small number of different substituents in a definite position often makes it impossible to derive a meaningful

Hansch correlation. On the other hand, a Hansch equation obtained from de novo group contributions which include a number of single point determinations is not very reliable because group contributions derived from a greater number of compounds and group contributions including the experimental error of a single compound have equal weights; according to eq 8 the mixed approach should be used in such cases to derive the correct Hansch correlation from the original log 1/C values.

The second application of eq 8 and 9 is the use of Free-Wilson like dummy variables (indicator variables<sup>22</sup>) in Hansch analysis. Different kinds of dummy variables have been used in Hansch and Free-Wilson analyses: (a) dummy variables to account for specific structural properties (e.g., steric hindrance or conformational changes), (b) dummy variables to account for specific interactions between substituents (e.g., hydrogen bonding), and (c) dummy variables to account for all changes in polar, electronic, and steric contributions if a substituent X is replaced by a substituent Y. A list of different dummy variables used in the literature is given in Table II.

While most authors used dummy variables without discussing a rational basis for their use, Hansch and Yoshimoto<sup>22</sup> tried to give an explanation; they interpreted the function of dummy variables (indicator variables) in Hansch analysis as "the extrathermodynamic approach assisted by what is sometimes termed the Free-Wilson method" and defined the Free-Wilson approach as a "quantitative structure-activity relationship formulated in terms of dummy variables" (it must be noted that a short discussion of the usefulness of dummy variables to combine different sets of compounds in a single Hansch equation was given by Franke and Oehme<sup>35</sup>).

Equations 8 and 9 support the view of Hansch and Yoshimoto; a dummy variable which accounts for all changes in polar, electronic, and steric contributions if a substituent X is replaced by a substituent Y (type c) is unequivocally a Free-Wilson group contribution of the substituent Y, based on  $a_X = 0.00$ ; the use of such dummy variables in a Hansch equation is an implied application of the mixed approach.

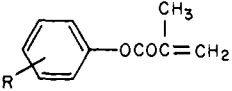
The situation is more difficult if dummy variables of type a are considered; these dummy variables may be interpreted as Free-Wilson group contributions from

Table II. Dummy Variables (Indicator Variables) Used in Hansch and Free-Wilson Analyses

Dummy variable	Used to account for	Type of dummy variable <sup>a</sup>	Ref
$n_H$	Number of hydrogens on protonated amine	a	23
$D$	Differences between phenyl and benzoyl derivatives	a	24
$X$	Stereoelectronic differences between meta and para substituents	a	25
$\delta$	Differences between phenyl and 2-thienyl derivatives	c	26
$X^*$	Hydrogen bonding between OH and $OCH_3$	b	5
$Ac$	Differences between OH and $OCOCH_3$	c	27
$T$	Differences between cis and trans isomers	a	27
$Et$	Steric and/or electronic differences between NMe and NET	a	27
$D$	Deviations found for $m-OCH_3$	a	28
$D(4'')$	Nonpolar differences between OH and OCOR	a	29
$D(11)$	Nonpolar differences between OH and OCOR	a	29
$D(B)$	Nonpolar differences between H and OH	a	29
$NMe$	Differences between NH and $NCH_3$	c	30
$I$	Nonpolar differences between indan and tetralin system	a	31
$X$	Different position of substituent X	a	31
$D-1$	Bridge units between two aromatic rings, including the second ring	c	22
$D-2$	Nonpolar differences between benzene and pyridine derivatives	a	22
$D-3$	Special activating effect of $-NHCOXC_6H_5$ in meta position	a	22
$D$	Activating effect of an oxygen substituent in para position	a	32
$x$	Differences in conformation between anilides and benzylamides	a	33
$D$	Nonpolar differences between mono- and disubstituted amides	a	34
$A$	Specific effects of a piperidine ring in substituted hydrazides	a	34

<sup>a</sup> See text.

Table III. Activity of Halogenated Phenylmethacrylates against Microorganisms



Compd no.	R	Log P <sup>a</sup>	Log 1/C values					
			IIIa <sup>b</sup>	IIIb <sup>c</sup>	IIIc <sup>d</sup>	IIId <sup>e</sup>	IIIe <sup>f</sup>	IIIg <sup>g</sup>
1	H	1.99	2.89	2.89	2.89	2.89	2.89	2.89
2	2-Cl	2.58	3.08	3.08	3.08	3.08	3.08	3.28
3	4-Cl	2.69	3.12	3.25	3.25	3.25	3.25	3.25
4	3-Cl	2.75	2.91	2.77	2.91	2.91	3.12	3.12
5	2,4-Cl <sub>2</sub>	3.28	3.49	3.49	3.80	3.49	3.49	3.49
6	2,4,6-Cl <sub>3</sub>	3.87	3.76	3.76	3.76	3.76	3.24	3.46
7	2,4,5-Cl <sub>3</sub>	4.04	3.84	3.84	3.84	3.84	3.54	3.54
8	2,3,4,6-Cl <sub>4</sub>	4.63	4.97	3.97	4.67	4.45	3.32	3.45
9	Cl <sub>5</sub>	5.39	5.19	4.71	5.00	5.00	3.71	3.35
10	Br <sub>5</sub>	6.39	5.32	3.64	5.32	3.85	3.64	3.35

<sup>a</sup> See ref 37. <sup>b</sup> vs. *S. aureus*. <sup>c</sup> vs. *S. faecalis*. <sup>d</sup> vs. *B. subtilis*. <sup>e</sup> vs. *B. cereus*. <sup>f</sup> vs. *Sarcina lutea*. <sup>g</sup> vs. *H. anomala*.<sup>37</sup>

Table IV. Matrices Used for Free-Wilson and Mixed Analyses

Compd no.	R	[Cl]	[o-Cl]	[m-Cl]	[p-Cl]	[o'-Cl]	$\pi^{15}$	$E_s^{\text{ortho}^a}$
1	H	0	0	0	0	0	0.00	1.24
2	2-Cl	1	1	0	0	0	0.71	1.24
3	4-Cl	1	0	0	1	0	0.71	1.24
4	3-Cl	1	0	1	0	0	0.71	1.24
5	2,4-Cl <sub>2</sub>	2	1	0	1	0	1.42	1.24
6	2,4,6-Cl <sub>3</sub>	3	2	0	1	1	2.13	0.27
7	2,4,5-Cl <sub>3</sub>	3	1	1	1	0	2.13	1.24
8	2,3,4,6-Cl <sub>4</sub>	4	2	1	1	1	2.84	0.27
9	Cl <sub>5</sub>	5	2	2	1	1	3.55	0.27

<sup>a</sup> Reference 38.

which one or more structural parameters have been "extracted".

The mixed approach (eq 7, 8, or 9) combines the linear free energy related Hansch model and the modified Free-Wilson model<sup>5,8</sup> and makes use of the specific advantages of each model; the following examples are given to illustrate the application of the mixed approach.

**Example 1** is an application of the Free-Wilson approach and the mixed approach (in form of eq 7) to describe the activity of chlorinated phenyl methacrylates against several microorganisms. Hansch et al.<sup>36,37</sup> derived for compounds 1-10 (for structures, log P values, and log 1/C values, see Table III) the following equations (eq 10-15).

vs. *S. aureus* (IIIa):

$$\log 1/C = 0.668 \log P + 1.342 \quad (10)$$

$n = 10; r = 0.966; s = 0.262$

vs. *S. faecalis* (IIIb):

$$\log 1/C = -0.125 (\log P)^2 + 1.359 \log P + 0.415 \quad (11)$$

$n = 10; r = 0.861; s = 0.334$

vs. *B. subtilis* (IIIc):

$$\log 1/C = 0.617 \log P + 1.530 \quad (12)$$

$n = 10; r = 0.976; s = 0.204$

vs. *B. cereus* (IIId):

$$\log 1/C = 0.400 \log P + 2.144 \quad (13)$$

$n = 10; r = 0.815; s = 0.420$

vs. *Sarcina lutea* (IIIe):

$$\log 1/C = 0.161 \log P + 2.721 \quad (14)$$

$n = 10; r = 0.849; s = 0.148$

vs. *H. anomala* (IIIg):

$$\log 1/C = -0.102 (\log P)^2 + 1.234 \log P - 0.880 + 0.878 \quad (15)$$

$n = 10; r = 0.958; s = 0.069$

While eq 10, 12, and 15 give good correlations of observed and calculated log 1/C values, the correlations found with eq 11, 13, and 14 are not so good. In order to explain the reasons responsible for these lower correlation coefficients some Free-Wilson analyses and mixed analyses were run (the matrices used for both approaches are given in Table IV; the pentabromo compound was excluded from all analyses to avoid single point determinations).

Based on the assumption that all chlorine atoms give comparable group contributions the following Free-Wilson correlations using [Cl] as single parameter were derived.

vs. *S. aureus* (IIIa):

$$\log 1/C = 0.503 (\pm 0.130) [\text{Cl}] + 2.578 \quad (16)$$

$n = 9; r = 0.960; s = 0.256; F = 83.06; p < 0.001$

vs. *S. faecalis* (IIIb):

$$\log 1/C = 0.359 (\pm 0.093) [\text{Cl}] + 2.732 \quad (17)$$

$n = 9; r = 0.961; s = 0.182; F = 84.03; p < 0.001$

vs. *B. subtilis* (IIIc):

$$\log 1/C = 0.443 (\pm 0.107) [\text{Cl}] + 2.705 \quad (18)$$

$n = 9; r = 0.966; s = 0.209; F = 96.65; p < 0.001$

vs. *B. cereus* (IIIId):

$$\log 1/C = 0.428 (\pm 0.086) [\text{Cl}] + 2.680 \quad (19)$$

$n = 9; r = 0.976; s = 0.169; F = 138.65; p < 0.001$

vs. *Sarcina lutea* (IIIe):

$$\log 1/C = 0.127 (\pm 0.079) [\text{Cl}] + 3.012 \quad (20)$$

$n = 9; r = 0.819; s = 0.116; F = 14.25; p < 0.01$

vs. *H. anomala* (IIIIf):

$$\log 1/C = 0.088 (\pm 0.082) [\text{Cl}] + 3.120 \quad (21)$$

$n = 9; r = 0.693; s = 0.160; F = 6.46; p < 0.05$

Equations 16-19 show good correlations between observed and calculated log 1/C values; there is no evidence for a parabolic dependence of log 1/C on lipophilic character. The good correlations found with eq 17 and 19 are rather surprising because eq 11 and 13 (including the pentabromo compound) gave smaller correlations. However, these smaller correlations are caused only by the pentabromo compound, as can be seen from eq 22 and 23, where the pentabromo compound was excluded.

vs. *S. faecalis* (IIIb):

$$\log 1/C = 0.535 (\pm 0.142) \log P + 1.674 \quad (22)$$

$n = 9; r = 0.959; s = 0.186; F = 79.52; p < 0.001$

vs. *B. cereus* (IIIId):

$$\log 1/C = 0.640 (\pm 0.126) \log P + 1.412 \quad (23)$$

$n = 9; r = 0.977; s = 0.165; F = 144.96; p < 0.001$

Whether the deviations found for the pentabromo compound are due to experimental error or to parabolic dependence of log 1/C values on lipophilic character or to specific changes in activity going from chlorine to bromine substitution cannot be decided with the available data. The significance of eq 22 and 23 is supported by the fact that eq 10, 12, 22, and 23 have similar coefficients; indeed log 1/C values IIIa, IIIb, IIIc, and IIIId (excluding the pentabromo compound) can be described by a single Hansch equation.

$$\log 1/C = 0.646 (\pm 0.071) \log P + 1.394 \quad (24)$$

$n = 36; r = 0.954; s = 0.217; F = 341.80; p < 0.001$

Equations 20 and 21 have low correlation coefficients; therefore a mixed approach (eq 7) was run for log 1/C values IIIe and IIIIf, using  $\pi^2$  and [Cl] as parameters.

vs. *Sarcina lutea* (IIIe):

$$\log 1/C = -0.031 (\pm 0.120) \pi^2 + 0.204 (\pm 0.314) [\text{Cl}] + 2.954 \quad (25)$$

$n = 9; r = 0.831; s = 0.163; F = 6.70; p < 0.05$

vs. *H. anomala* (IIIIf):

$$\log 1/C = -0.119 (\pm 0.046) \pi^2 + 0.387 (\pm 0.120) [\text{Cl}] + 2.895 \quad (26)$$

$n = 9; r = 0.965; s = 0.063; F = 41.16; p < 0.001$

Log 1/C values IIIIf (activity vs. *H. anomala*) show a significant parabolic dependence on lipophilic character ( $t_{\pi^2} = 6.32; p < 0.001$ ); a comparison of the correlation coefficients and the  $F$  values of eq 21 and 26 illustrates the superiority of eq 26 to describe the relationship between

chemical structure and biological activity for log 1/C values IIIIf. Equation 26 corresponds to the parabolic eq 15 (no  $\pi_{\text{max}}$  can be calculated from eq 26 because the first-order  $\pi$  term is "hidden" in the Free-Wilson parameter).

For log 1/C values IIIe (activity vs. *Sarcina lutea*) no satisfactory correlation can be obtained either by Hansch analysis (eq 14) or Free-Wilson analysis (eq 20) or mixed analysis (eq 25). Therefore a second Free-Wilson analysis was run with [*o*-Cl], [*m*-Cl], and [*p*-Cl] as parameters (see Table IV) to prove the assumption that all chlorine atoms have equal group contributions; additional  $\pi^2$  term did not

$$\log 1/C = -0.001 (\pm 0.185) [o\text{-Cl}] + 0.168 (\pm 0.185) [m\text{-Cl}] + 0.340 (\pm 0.303) [p\text{-Cl}] + 2.974 \quad (27)$$

$n = 9; r = 0.906; s = 0.136; F = 7.64; p < 0.05$

improve the correlation. Although eq 27 gives a somewhat better correlation than eq 14, 20, or 25, the low coefficient found for the [*o*-Cl] term indicates that something may be "wrong" with the ortho position.

Next a Free-Wilson analysis was run to prove the equivalence of *o*- and *o'*-chlorine atoms; the total number of chlorine atoms [Cl] and a dummy variable [*o'*-Cl] (see Table IV), accounting for ortho,ortho'-disubstitution, were used as parameters.

$$\log 1/C = 0.231 (\pm 0.082) [\text{Cl}] - 0.421 (\pm 0.270) [o'\text{-Cl}] + 2.921 \quad (28)$$

$n = 9; r = 0.951; s = 0.091; F = 28.17; p < 0.001$

The [*o'*-Cl] term is a dummy variable of type a; its high negative value indicates that ortho,ortho'-disubstitution causes a significant reduction of biological activity. Since this reduction of activity may be due to a negative steric contribution of the second *o*-chlorine atom, log  $P$  and  $E_{s,\text{ortho}'}$  (see Table IV) were used to calculate eq 29. Of

$$\log 1/C = 0.328 (\pm 0.123) \log P + 0.388 (\pm 0.278) E_{s,\text{ortho}'} + 1.800 \quad (29)$$

$n = 9; r = 0.946; s = 0.095; F = 25.63; p < 0.005$

course eq 28 and 29 may be chance correlations;<sup>39</sup> although the  $E_{s,\text{ortho}'}$  term is statistically significant ( $t_{E_{s,\text{ortho}'}} = 3.42; p < 0.05$ ) there is no further evidence that these compounds exert their action vs. *Sarcina lutea* by a specific mechanism. Some evidence for the biological significance of eq 29 may be derived from the fact that the biological activity of the pentabromo compound ( $E_{s,\text{ortho}'} = 0.08$ ) is also well predicted by eq 29 (log 1/C obsd = 3.64; log 1/C calcd = 3.93).

The purpose of this example was to demonstrate the step-by-step analysis of biological activity data by Free-Wilson analysis, mixed analysis, and finally Hansch analysis with a limited number of compounds.

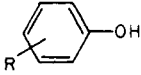
**Example 2** demonstrates the reliability of the mixed approach (eq 7) in detecting a parabolic dependence of log 1/C values on lipophilic character also in cases where a nonparabolic Free-Wilson analysis gives a very high correlation of observed and calculated log 1/C values.

Hansch and Lien<sup>37</sup> derived for the inhibitory activity of substituted phenols (for structures see Table V) against *Aspergillus niger* eq 30. Although the (log  $P$ )<sup>2</sup> term is

$$\log 1/C = -0.190 (\pm 0.085) (\log P)^2 + 1.859 (\pm 0.558) \log P + 0.627 (\pm 0.395) \sigma - 0.092 \quad (30)$$

$n = 18; r = 0.975; s = 0.160; F = 90.35; p < 0.001$

Table V. Growth Inhibitory Activity of Substituted Phenols vs. *Aspergillus niger*. Structures, Structural Parameters and Log 1/C Values

Compd no.	R				$\sigma^a$	Log 1/C obsd <sup>a</sup>
		Log $P^a$	$\pi^b$			
1	H	1.46	0.00	0.00	2.35	
2	4-Cl	2.39	0.93 <sup>c</sup>	0.23	3.35	
3	2-Me	1.96	0.56	-0.14	2.70	
4	2-Me, 4-Cl	2.89	1.49	0.09	3.70	
5	3-Me	2.02	0.56	-0.07	2.68	
6	3-Me, 4-Cl	2.95	1.49	0.16	3.70	
7	2,6-Me <sub>2</sub>	2.46	1.12	-0.28	3.35	
8	2,6-Me <sub>2</sub> , 4-Cl	3.39	2.05	-0.05	4.40	
9	3,5-Me <sub>2</sub>	2.58	1.12	-0.14	3.26	
10	3,5-Me <sub>2</sub> , 4-Cl	3.51	2.05	0.09	4.22	
11	2- <i>i</i> -Pr	2.76	1.53	-0.23	3.35	
12	2- <i>i</i> -Pr, 4-Cl	3.69	2.46	0.00	4.30	
13	2- <i>t</i> -Bu, 5-Me	3.70	2.54	-0.59	3.70	
14	2- <i>t</i> -Bu, 4-Cl, 5-Me	4.63	3.47	-0.36	4.30	
15	2-Cyclohexyl	3.97	2.51	-0.23	4.00	
16	2-Cyclohexyl, 4-Cl	4.90	3.44	0.00	4.40	
17	2-Phenyl	3.59	1.96	0.00	4.10	
18	2-Phenyl, 4-Cl	4.52	2.89	0.23	4.52	

<sup>a</sup> Reference 37. <sup>b</sup> From substituted benzenes, ref 15, unless otherwise indicated. <sup>c</sup> From the phenol system, ref 14.

statistically significant ( $t_{(\log P)^2} = 4.78$ ;  $p < 0.001$ ) and its numerical contribution to log 1/C values is high, the biological activity of this group of compounds can also be described by a nonparabolic relationship (eq 31).

$$\log 1/C = 0.619 (\pm 0.146) \log P + 1.715 \quad (31)$$

$n = 18$ ;  $r = 0.914$ ;  $s = 0.275$ ;  $F = 80.79$ ;  $p < 0.001$

Therefore a high correlation coefficient is obtained, due to the greater number of parameters, from a nonparabolic Free-Wilson analysis (matrix see Table VI).

Table VI. Matrix Used for Free-Wilson and Mixed Analysis

Compd no.	[ <i>p</i> -Cl]	[ <i>o</i> -Me]	[ <i>o</i> - <i>i</i> -Pr]	[ <i>o</i> - <i>t</i> -Bu]	[ <i>o</i> -C <sub>6</sub> H <sub>11</sub> ]	[ <i>o</i> -Phe]	[ <i>m</i> -Me]	$\pi^{2a}$
1								0.00
2	1							0.86
3		1						0.31
4	1	1						2.22
5							1	0.31
6	1						1	2.22
7		2						1.25
8	1	2						4.20
9							2	1.25
10	1						2	4.20
11			1					2.34
12	1		1					6.05
13				1			1	6.45
14	1			1			1	12.04
15					1			6.30
16	1				1			11.83
17						1		3.84
18	1					1		8.35

<sup>a</sup>  $\pi$  values of Table V were used to calculate  $\pi^2$ .

$$\log 1/C = 0.822 [p\text{-Cl}] + 0.526 [o\text{-Me}] + 1.051 [o\text{-}i\text{-Pr}] + 0.757 [o\text{-}t\text{-Bu}] + 1.426 [o\text{-}C_6H_{11}] + 1.536 [o\text{-Phe}] + 0.470 [m\text{-Me}] + 2.363 \quad (32)$$

$n = 18$ ;  $r = 0.977$ ;  $s = 0.183$ ;  $F = 29.93$ ;  $p < 0.001$

A mixed approach with an additional  $\pi^2$  term (matrix see Table VI) gave eq 33. The  $\pi^2$  term in eq 33 is sta-

$$\log 1/C = -0.124 (\pm 0.091) \pi^2 + 1.235 [p\text{-Cl}] + 0.674 [o\text{-Me}] + 1.541 [o\text{-}i\text{-Pr}] + 1.726 [o\text{-}t\text{-Bu}] + 2.522 [o\text{-}C_6H_{11}] + 2.263 [o\text{-Phe}] + 0.618 [m\text{-Me}] + 2.187 \quad (33)$$

$n = 18$ ;  $r = 0.989$ ;  $s = 0.134$ ;  $F = 50.10$ ;  $p < 0.001$

tistically significant ( $t_{\pi^2} = 3.11$ ;  $p < 0.05$ ); although there is no significant difference in the correlation coefficients obtained with eq 32 and 33, the higher  $F$  value of eq 33 supports the parabolic dependence of log 1/C values on lipophilic character.

A mixed approach with a different  $\pi$  value for the *p*-Cl substituent ( $\pi = 0.71^{15}$  instead of  $\pi = 0.93$ , see Table V) gave a comparable result ( $r = 0.988$ ;  $s = 0.138$ ;  $F = 47.14$ ;  $p < 0.001$ ;  $t_{\pi^2} = 2.93$ ;  $p < 0.05$ ), indicating that the selection of  $\pi$  values is not very important in this case.

**Example 3** is taken from a recent paper on structure-activity correlations among rifamycin B amides and hydrazides.<sup>34</sup> Only the substituted hydrazides will be considered in this example to demonstrate the usefulness of the mixed approach and to show some possible pitfalls in Hansch analysis of an inhomogeneous set of compounds.

Quinn et al.<sup>34</sup> derived for the antibacterial activity of substituted rifamycin B hydrazides vs. five different bacterial systems (structures and log 1/C values, see Table VII) structure-activity correlations but they stated that the large amount of collinearity between their structural parameters log  $P$ ,  $\sigma^*$ , and  $E_s$  makes it impossible to draw any firm conclusions about the real structure-activity relationships.

Compounds 52-66 and 68-71 were selected for Free-Wilson analysis and mixed analysis (for matrix see Table VIII); compounds 50, 51, and 67 were excluded to avoid single point determinations and compounds 72-75 were excluded because of nonexplicable deviations of log 1/C values for compounds 73 and 75 (compare ref 34).

The group contributions found by Free-Wilson analysis are given in the following equations (all group contribu-

Table VII. Structures and Antibacterial Activities of Substituted Rifamycin B Hydrazides vs. Several Bacterial Systems (All Values Taken from Ref 34)

Compd no.	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	Log 1/C values					Log P	σ*
				VIIa <sup>a</sup>	VIIb <sup>b</sup>	VIIc <sup>c</sup>	VIIId <sup>d</sup>	VIIe <sup>e</sup>		
50	H	H	H	6.98	6.04	7.11	5.40	6.71	-1.93	1.47
51	H	H	Phe	6.93	5.93	7.23	5.44	7.36	0.21	1.58
52	Me	Me	Me	8.13	7.21	8.61	6.65	6.65	-0.43	0.00
53	Me	Et	Et	8.62	7.62	8.62	6.96	6.35	0.57	-0.20
54	Me	Pr	Pr	8.94	7.76	8.64	6.98	6.37	1.57	-0.23
55	Me	Bu	Bu	8.87	7.95	8.65	7.30	7.00	2.57	-0.26
56	Et	Me	Me	8.62	7.44	8.62	6.62	6.62	0.07	-0.10
57	Et	Et	Et	8.93	7.63	8.63	7.63	6.76	1.07	-0.30
58	Et	Pr	Pr	8.87	7.77	8.87	7.29	6.69	2.07	-0.33
59	Et	Bu	Bu	8.80	7.96	8.96	7.31	7.00	3.07	-0.36
60	Pr	Me	Me	8.43	7.60	8.60	6.65	6.73	0.57	-0.12
61	Pr	Et	Et	8.94	7.63	8.64	7.64	6.76	1.57	-0.32
62	Pr	Pr	Pr	8.48	7.65	8.48	7.17	7.00	2.57	-0.34
63	Pr	Bu	Bu	8.49	7.66	8.67	7.49	7.01	3.57	-0.38
64	Bu	Me	Me	8.93	7.68	8.33	7.28	6.76	1.07	-0.13
65	Bu	Et	Et	8.64	7.94	7.94	7.64	6.69	2.07	-0.33
66	Bu	Pr	Pr	8.66	7.66	8.66	7.31	7.00	3.07	-0.36
67	Pentyl	Me	Me	8.64	7.94	8.09	7.28	6.68	1.57	-0.13
68	Me	-(CH <sub>2</sub> ) <sub>5</sub> -		8.63	7.63	7.93	6.98	6.98	0.67	-0.22
69	Et	-(CH <sub>2</sub> ) <sub>5</sub> -		8.64	7.64	7.94	6.98	7.28	1.17	-0.32
70	Pr	-(CH <sub>2</sub> ) <sub>5</sub> -		8.64	7.94	7.94	7.29	7.29	1.67	-0.34
71	Bu	-(CH <sub>2</sub> ) <sub>5</sub> -		8.65	7.95	8.11	7.30	7.00	2.17	-0.35
72	Me	-(CH <sub>2</sub> ) <sub>2</sub> O(CH <sub>2</sub> ) <sub>2</sub> -		8.23	7.68	7.76	6.36	6.68	-1.26	0.18
73	Et	-(CH <sub>2</sub> ) <sub>2</sub> O(CH <sub>2</sub> ) <sub>2</sub> -		7.64	6.68	8.86	6.06	6.37	-0.76	0.08
74	Pr	-(CH <sub>2</sub> ) <sub>2</sub> O(CH <sub>2</sub> ) <sub>2</sub> -		8.10	7.17	8.25	6.38	6.38	-0.26	0.06
75	Bu	-(CH <sub>2</sub> ) <sub>2</sub> O(CH <sub>2</sub> ) <sub>2</sub> -		7.65	6.70	7.65	5.78	5.78	0.24	0.05

<sup>a</sup> vs. *M. aureus*. <sup>b</sup> vs. *S. faecalis*. <sup>c</sup> vs. *S. hemolyticus*. <sup>d</sup> vs. *B. subtilis*. <sup>e</sup> vs. *M. tuberculosis*.

Table VIII. Matrix Used for Free-Wilson and Mixed Analysis (All Group Contributions Based on  $\alpha_{Me} = 0.00$ )

Compd no.	[Et]	[Pr]	[Bu]	[Pip]	(Log P) <sup>2a</sup>
52					0.18
53	2				0.32
54 <sup>b</sup>		2			2.46
55			2		6.60
56	1				0.00
57	3				1.14
58	1	2			4.28
59	1		2		9.42
60		1			0.32
61	2	1			2.46
62		3			6.60
63		1	2		12.74
64			1		1.14
65 <sup>c</sup>	2		1		4.28
66		2	1		9.42
68				1	0.45
69	1			1	1.37
70		1		1	2.79
71			1	1	4.71

<sup>a</sup> Log P values of Table VII were used to calculate (log P)<sup>2</sup>. <sup>b</sup> Compound not included in calculation of eq 38. <sup>c</sup> Compound not included in calculation of eq 36.

utions based on  $\alpha_{Me} = 0.00$ ).

vs. *M. aureus* (VIIa):

$$\log 1/C = 0.129 [\text{Et}] + 0.070 [\text{Pr}] + 0.100 [\text{Bu}] + 0.095 [\text{Pip}] + 8.471 \quad (34)$$

$n = 19; r = 0.493; s = 0.209; F = 1.12; p > 0.1$

vs. *S. faecalis* (VIIb):

$$\log 1/C = 0.098 [\text{Et}] + 0.101 [\text{Pr}] + 0.210 [\text{Bu}] + 0.296 [\text{Pip}] + 7.392 \quad (35)$$

$n = 19; r = 0.768; s = 0.140; F = 5.02; p = 0.01$

vs. *S. hemolyticus* (VIIc):

$$\log 1/C = 0.072 [\text{Et}] + 0.041 [\text{Pr}] + 0.094 [\text{Bu}] - 0.573 [\text{Pip}] + 8.501 \quad (36)$$

$n = 18; r = 0.930; s = 0.132; F = 20.77; p < 0.001$

vs. *B. subtilis* (VIIId):

$$\log 1/C = 0.300 [\text{Et}] + 0.209 [\text{Pr}] + 0.349 [\text{Bu}] + 0.347 [\text{Pip}] + 6.576 \quad (37)$$

$n = 19; r = 0.847; s = 0.194; F = 8.86; p < 0.001$

vs. *M. tuberculosis* (VIIe):

$$\log 1/C = 0.007 [\text{Et}] + 0.119 [\text{Pr}] + 0.160 [\text{Bu}] + 0.472 [\text{Pip}] + 6.594 \quad (38)$$

$n = 18; r = 0.861; s = 0.139; F = 9.27; p < 0.001$

A mixed analysis of log 1/C values VIIa and VIIb (for matrix see Table VIII) gave eq 39 and 40.

vs. *M. aureus* (VIIa):

$$\log 1/C = -0.137 (\pm 0.072) (\log P)^2 + 0.255 [\text{Et}] + 0.444 [\text{Pr}] + 0.770 [\text{Bu}] + 0.362 [\text{Pip}] + 8.231 \quad (39)$$

$n = 19; r = 0.821; s = 0.142; F = 5.37; p < 0.01;$   
 $t_{(\log P)^2} = 4.15; p < 0.005$

vs. *S. faecalis* (VIIb):

$$\log 1/C = -0.099 (\pm 0.042) (\log P)^2 + 0.189 [\text{Et}] + 0.371 [\text{Pr}] + 0.694 [\text{Bu}] + 0.489 [\text{Pip}] + 7.219 \quad (40)$$

$n = 19; r = 0.928; s = 0.084; F = 16.16; p < 0.001;$   
 $t_{(\log P)^2} = 5.05; p < 0.001$

The statistically significant  $(\log P)^2$  terms and the higher  $F$  values of eq 39 and 40, compared with eq 34 and 35, indicate a parabolic dependence of  $\log 1/C$  values on lipophilic character for the activity of the trisubstituted hydrazides vs. *M. aureus* and *S. faecalis*. No parabolic dependence could be detected for  $\log 1/C$  values VIIc, VIId, and VIIe (activity vs. *S. hemolyticus*, *B. subtilis*, and *M. tuberculosis*); all  $(\log P)^2$  terms were statistically not significant ( $p > 0.1$ ). A matrix using different Free-Wilson parameters for  $R_1$  and  $R_2/R_3$  gave only somewhat higher correlation coefficients than eq 34-40 but lower  $F$  values, indicating that the group contributions of the alkyl substituents are not significantly different in position  $R_1$  and positions  $R_2/R_3$ .

The fact that the biological activity of the trisubstituted hydrazides vs. *M. aureus* cannot be described by a nonparabolic equation (eq 34;  $p > 0.1$ ) was surprising because Quinn et al.<sup>34</sup> derived eq 41 for compounds 50-72 and 74,

$$\log 1/C = -0.96 (\pm 0.16) \sigma^* + 8.42 \quad (41)$$

$n = 24; r = 0.937; s = 0.189$

which is a nonparabolic equation with a significantly better correlation coefficient than eq 34 ( $r = 0.493$ ).

A Hansch analysis of compounds 52-66 and 68-71 (set of compounds used for Free-Wilson and mixed analysis) with  $\sigma^*$  as single parameter gave eq 42.

$$\log 1/C = -0.72 (\pm 0.93) \sigma^* + 8.49 \quad (42)$$

$n = 19; r = 0.370; s = 0.202; F = 2.70; p > 0.1$

If compounds 50 and 51 are included, eq 43 is obtained for compounds 50-66 and 68-71.

$$\log 1/C = -0.96 (\pm 0.16) \sigma^* + 8.43 \quad (43)$$

$n = 21; r = 0.941; s = 0.193; F = 147.32; p < 0.001$

The great differences found for the correlation coefficients and the  $F$  values of eq 42 and 43 are caused only by compounds 50 and 51; that means that eq 43 only can describe the differences in the biological activities between compounds 50 and 51 on one hand and compounds 52-66 and 68-71 on the other hand.

According to this a correlation can be established for compounds 50-66 and 68-71 using an indicator variable  $D$  as unique parameter ( $D = 1$  for all trisubstituted hydrazides;  $D = 0$  for compounds 50 and 51).

$$\log 1/C = 1.72 (\pm 0.32) D + 6.96 \quad (44)$$

$n = 21; r = 0.933; s = 0.206; F = 127.02; p < 0.001$

From eq 44 it is evident that eq 41, 42, and 43 are without any significance in predicting differences in biological activity among the trisubstituted hydrazides 52-66 and 68-71; eq 41 and 43 are chance correlations.<sup>39</sup> This example demonstrates that care must be taken in the selection of compounds for Hansch analysis; compounds with different structures and very different activities (such as compounds 50 and 51) should be excluded if they have such a great influence on the overall result (note that Tute<sup>40</sup> came to a similar conclusion discussing a theoretical example, ref 40, pp 41-42).

Including only trisubstituted hydrazides the following equation can be derived for compounds 52-66 and 68-71.

$$\log 1/C = -0.111 (\pm 0.062) (\log P)^2 + 0.411$$

$$(\pm 0.213) \log P + 8.427 \quad (45)$$

$n = 19; r = 0.715; s = 0.157; F = 8.38; p < 0.005$

No equation with a statistically significant  $\sigma^*$  term could be derived; a correlation using  $(\log P)^2$ ,  $\log P$ , and  $E_s$  ( $n = 19; r = 0.797; s = 0.140; F = 8.73; p < 0.005$ ) is not presented here because it may be a chance correlation due to the collinearity of  $\log P$  and  $E_s$  (see ref 34). The low correlation coefficient of eq 45 is not surprising if the small variance of  $\log 1/C$  values (0.8 log units) and the usual experimental error (ca. 0.2 log units) are taken into account. Of course eq 45 and 39 are valid only for trisubstituted hydrazides.

Because of the collinearity problem<sup>34</sup> no attempts were made to derive Hansch correlations for the biological activity of the trisubstituted hydrazides vs. *S. faecalis*, *S. hemolyticus*, *B. subtilis*, and *M. tuberculosis* (it should be noted that the use of a dummy variable A for the piperidine substituent<sup>34</sup> is confirmed by the group contributions found with eq 36 and 38).

Example 4 illustrates the application of the mixed approach in the form of eq 8 and 9.

$$\log 1/C = \sum_i a_i + \sum_j k_j \phi_j + k' \quad (8)$$

$$\log 1/C = k_1 \pi^2 + \sum_i a_i + \sum_j k_j \phi_j + k' \quad (9)$$

Hansch and Clayton<sup>9</sup> derived for the bactericidal activity of substituted benzylammonium compounds against *S. aureus* (for structures and  $\log 1/C$  values see Table IX; for  $\log P$  values see ref 9) eq 46.

$$\log 1/C = -0.17 (\pm 0.03) (\log P)^2 + 0.87 (\pm 0.14)$$

$$\log P + 2.93 \quad (46)$$

$n = 45; r = 0.884; s = 0.306$

In order to find a better correlation between structure and biological activity two mixed analyses were run (corresponding to eq 8 and 9) with a Free-Wilson part  $\sum a_i$  for the aromatic substituents  $X_i$  (2-Cl, 3-Cl, 4-Cl, 4-NO<sub>2</sub>, -OCH<sub>3</sub> and 3,4-OCH<sub>2</sub>O-), a Hansch part  $k_2 \pi_R$  for the aliphatic side chains R (C<sub>8</sub>, C<sub>10</sub>, C<sub>12</sub>, C<sub>14</sub>, C<sub>16</sub>, C<sub>18</sub>), and an additional term  $k_1 \pi^2$  to account for parabolic dependence of  $\log 1/C$  values on lipophilic character (for matrix see Table X; compound 45 was excluded to avoid single point determinations).

The aliphatic side chains are described by  $\pi_R$  because there is a high degree of probability that the change in lipophilic character is the only physical parameter influencing the biological activity in going from a C<sub>8</sub> side chain to longer side chains; the aromatic substituents  $X_i$  are described by Free-Wilson parameters in order to check the possibility whether electronic and/or steric effects of these substituents may contribute to biological activity.

A nonparabolic mixed approach (eq 8) gave no statistically significant correlation (for matrix see Table X).

$$\log 1/C = 0.150 (\pm 0.394) \pi_R - 0.025 [2-Cl] +$$

$$0.095 [4-Cl] + 0.008 [4-NO_2] - 0.004$$

$$[3-Cl] + 0.079 [OCH_3] - 0.047 [OCH_2O] +$$

$$3.136 \quad (47)$$

$n = 44; r = 0.407; s = 0.632; F = 1.02; p > 0.1$

This is a further confirmation for the nonvalidity of the Free-Wilson approach in the case of significant parabolic dependence of  $\log 1/C$  values on lipophilic character<sup>8,10</sup> (eq 47 is indeed a Free-Wilson type equation because the  $\pi_R$  term may also be interpreted as a Free-Wilson parameter of a -CH<sub>2</sub>CH<sub>2</sub> group, based on the C<sub>8</sub> compound).

Table IX. Bactericidal Activity of Substituted Benzylammonium Compounds vs. *S. aureus*. Structures and Log 1/C Values<sup>a</sup>

Compd no.	X	R	Log 1/C obsd	Compd no.	X	R	Log 1/C obsd
1	H	C <sub>10</sub>	2.79	24	2,4-(Cl) <sub>2</sub>	C <sub>8</sub>	2.63
2	H	C <sub>12</sub>	3.74	25	2,4-(Cl) <sub>2</sub>	C <sub>10</sub>	3.65
3	H	C <sub>14</sub>	4.17	26	2,4-(Cl) <sub>2</sub>	C <sub>12</sub>	4.25
4	H	C <sub>16</sub>	3.92	27	2,4-(Cl) <sub>2</sub>	C <sub>14</sub>	4.28
5	H	C <sub>18</sub>	3.34	28	2,4-(Cl) <sub>2</sub>	C <sub>16</sub>	3.41
6	2-Cl	C <sub>8</sub>	2.02	29	2,4-(Cl) <sub>2</sub>	C <sub>18</sub>	3.30
7	2-Cl	C <sub>10</sub>	3.41	30	3,4-(Cl) <sub>2</sub>	C <sub>8</sub>	2.92
8	2-Cl	C <sub>12</sub>	4.14	31	3,4-(Cl) <sub>2</sub>	C <sub>10</sub>	3.79
9	2-Cl	C <sub>14</sub>	4.14	32	3,4-(Cl) <sub>2</sub>	C <sub>12</sub>	4.36
10	2-Cl	C <sub>16</sub>	3.74	33	3,4-(Cl) <sub>2</sub>	C <sub>14</sub>	4.04
11	2-Cl	C <sub>18</sub>	3.43	34	3,4-(Cl) <sub>2</sub>	C <sub>16</sub>	3.39
12	4-Cl	C <sub>8</sub>	2.60	35	3,4-(Cl) <sub>2</sub>	C <sub>18</sub>	3.11
13	4-Cl	C <sub>10</sub>	3.60	36	3,4-(OCH <sub>3</sub> ) <sub>2</sub>	C <sub>12</sub>	3.71
14	4-Cl	C <sub>12</sub>	4.04	37	3,4-(OCH <sub>3</sub> ) <sub>2</sub>	C <sub>14</sub>	4.11
15	4-Cl	C <sub>14</sub>	4.34	38	3,4-(OCH <sub>3</sub> ) <sub>2</sub>	C <sub>16</sub>	3.41
16	4-Cl	C <sub>16</sub>	3.85	39	3,4-OCH <sub>2</sub> O-	C <sub>8</sub>	2.04
17	4-Cl	C <sub>18</sub>	3.17	40	3,4-OCH <sub>2</sub> O-	C <sub>10</sub>	3.08
18	4-NO <sub>2</sub>	C <sub>8</sub>	2.11	41	3,4-OCH <sub>2</sub> O-	C <sub>12</sub>	4.00
19	4-NO <sub>2</sub>	C <sub>10</sub>	3.11	42	3,4-OCH <sub>2</sub> O-	C <sub>14</sub>	4.23
20	4-NO <sub>2</sub>	C <sub>12</sub>	4.20	43	3,4-OCH <sub>2</sub> O-	C <sub>16</sub>	4.20
21	4-NO <sub>2</sub>	C <sub>14</sub>	4.23	44	3,4-OCH <sub>2</sub> O-	C <sub>18</sub>	3.23
22	4-NO <sub>2</sub>	C <sub>16</sub>	4.00	45	2-OH, 5-NO <sub>2</sub>	C <sub>12</sub>	2.70
23	4-NO <sub>2</sub>	C <sub>18</sub>	3.46				

The parabolic mixed approach (eq 9) gives a statistically significant correlation (eq 48); the values found for the

$$\log 1/C = -0.180 (\pm 0.030) \pi^2 + 1.293 (\pm 0.198) \pi_R + 0.760 [2\text{-Cl}] + 1.004 [4\text{-Cl}] + 0.449 [4\text{-NO}_2] + 0.995 [3\text{-Cl}] - 0.010 [\text{OCH}_3] + 0.123 [\text{OCH}_2\text{O}] + 1.709 \quad (48)$$

$n = 44; r = 0.916; s = 0.281; F = 22.93; p < 0.001$

individual group contributions are very different from those of eq 47.

The  $\pi^2$  term of eq 48 is statistically significant ( $t_{\pi^2} = 12.14; p < 0.001$ ).

A parabolic Hansch analysis using the same  $\pi$  values (from the phenoxyacetic acid system,<sup>14</sup> compare Tables X and XI) gives eq 49.

$$\log 1/C = -0.174 (\pm 0.028) \pi^2 + 1.253 (\pm 0.186) \pi + 1.836 \quad (49)$$

$n = 44; r = 0.907; s = 0.273; F = 95.52; p < 0.001$

The identity of the correlation coefficients found with eq 48 and 49 ( $r = 0.916$  vs.  $r = 0.907$ ) indicates that electronic or steric effects of the aromatic substituents have no influence on the biological activity.

An interesting effect is observed if different  $\pi$  values are used for the Hansch analyses;  $\pi$  values from the benzene system<sup>15</sup> ( $\pi$  values II of Table XI) give a somewhat smaller correlation (eq 50), arbitrarily chosen  $\pi$  values ( $\pi$  values III of Table XI) give a better correlation (eq 51) than the  $\pi$  values used for the calculation of eq 49.

$\pi$  values II (see Table XI)

$$\log 1/C = -0.153 (\pm 0.031) \pi^2 + 1.099 (\pm 0.202) \pi + 2.081 \quad (50)$$

$n = 44; r = 0.868; s = 0.323; F = 62.33; p < 0.001$

$\pi$  values III (see Table XI)

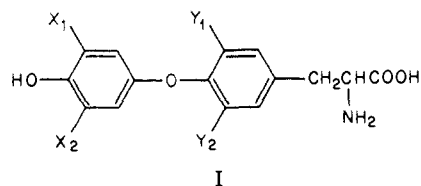
$$\log 1/C = -0.209 (\pm 0.023) \pi^2 + 1.340 (\pm 0.137) \pi + 2.011 \quad (51)$$

$n = 44; r = 0.952; s = 0.198; F = 199.71; p < 0.001$

From the different correlations found with eq 49, 50, and 51 one can see that the selection of  $\pi$  values may be critical in such cases where the influence of the  $\pi^2$  term is significant. In the given example the variance of the  $\pi$  values due to the chain length of the aliphatic residue is 5 log units; therefore the decision to use  $\pi = 0.24$  or  $\pi = -0.28$  for the *p*-NO<sub>2</sub> group (see Table XI) has great influence on the  $\pi^2$  values of the higher members.

The better fitting of the arbitrarily chosen  $\pi$  values ( $\pi$  values III, Table XI) indicates that nonadditivity due to chain folding<sup>13,17</sup> may be the reason for the smaller correlations found with eq 49 and 50 (note that the arbitrarily chosen  $\pi$  values are significantly smaller than the other  $\pi$  values).

**Example 5** is a further reinvestigation<sup>8</sup> of the thyroxine-like activity of thyronine derivatives I with the mixed approach (eq 8). It is a fascinating problem to



establish structure-activity relationships for this group of compounds for three reasons.

Firstly, a specific receptor with a definite structure and definite properties can be assumed for the site of action of a hormone and its analogues; therefore not only lipophilic character but also electronic and steric properties might have an influence on the biological activities of these compounds. Secondly, a large number of structurally similar compounds have been synthesized and tested, some



Table X. Matrix Used for the Mixed Analyses (Eq 47 and 48)

Compd no.	$\pi_R^a$	[2-Cl]	[4-Cl]	[4-NO <sub>2</sub> ]	[3-Cl]	[OCH <sub>3</sub> ]	[OCH <sub>2</sub> O]	$\pi^{2b}$
1	1							1.00
2	2							4.00
3	3							9.00
4	4							16.00
5	5							25.00
6	0	1						0.35
7	1	1						2.53
8	2	1						6.71
9	3	1						12.89
10	4	1						21.07
11	5	1						31.25
12	0		1					0.49
13	1		1					2.89
14	2		1					7.29
15	3		1					13.69
16	4		1					22.09
17	5		1					32.49
18	0			1				0.06
19	1			1				1.54
20	2			1				5.02
21	3			1				10.50
22	4			1				17.98
23	5			1				27.46
24	0	1	1					1.66
25	1	1	1					5.24
26	2	1	1					10.82
27	3	1	1					18.40
28	4	1	1					27.98
29	5	1	1					39.56
30	0		1		1			2.13
31	1		1		1			6.05
32	2		1		1			11.97
33	3		1		1			19.89
34	4		1		1			29.81
35	5		1		1			41.73
36	2					2		4.33
37	3					2		9.49
38	4					2		16.65
39	0						1	0.00
40	1						1	0.90
41	2						1	3.80
42	3						1	8.70
43	4						1	15.60
44	5						1	24.50

<sup>a</sup> The C<sub>8</sub> compound was taken as basis compound ( $\pi = 0.00$ ) and 0.50 was added for each additional -CH<sub>2</sub> group. <sup>b</sup>  $\pi = \pi_X + \pi_R$ ;  $\pi$  values from the phenoxyacetic acid system<sup>14</sup> ( $\pi$  values I of Table XI, see below) were used to calculate the  $\pi^2$  values.

Table XI.  $\pi$  Values Used for Calculation of Eq 48-51 (in All Cases  $\pi = 0.50$  was Used for the Aliphatic -CH<sub>2</sub> Group)

Substituent	$\pi$ values		
	I <sup>a</sup>	II <sup>b</sup>	III <sup>c</sup>
2-Cl	0.59	0.71	0.2
3-Cl	0.76	0.71	0.4
4-Cl	0.70	0.71	0.4
4-NO <sub>2</sub>	0.24	-0.28	0.0
3,4-(OCH <sub>3</sub> ) <sub>2</sub>	0.08	-0.04	0.0
3,4-OCH <sub>2</sub> O-	-0.05 <sup>b</sup>	-0.05	0.0

<sup>a</sup>  $\pi$  values from the phenoxyacetic acid system<sup>14</sup> unless otherwise indicated (used for eq 48 and 49). <sup>b</sup>  $\pi$  values from the benzene system<sup>15</sup> (used for eq 50). <sup>c</sup> Arbitrarily chosen  $\pi$  values to get a high correlation between observed and calculated log 1/C values (used for eq 51).

of them being as active as the natural hormones, so that enough data are available for the derivation of quantitative structure-activity correlations. Thirdly, many qualitative structure-activity relationships have been discussed for this group of compounds<sup>41-46</sup> and also some attempts were made to establish quantitative structure-activity correlations<sup>1,8,47-49</sup> but until today no quantitative correlation has been given which is valid for halogen- and alkyl-substituted thyronines.

In the preceding paper<sup>8</sup> a Hansch equation with an additional dummy variable  $D$  was derived for 13 halogen-substituted thyronine derivatives (eq 52, for

$$\log A = 2.54 (\pm 0.71) \pi_x + 2.91 (\pm 0.83) E_{s,x_1} - 0.35 (\pm 0.30) D - 3.72 \quad (52)$$

$n = 13; r = 0.942; s = 0.408; F = 23.45; p < 0.001$

structures and structural parameters see ref 8). The original log  $A$  values used by Bruice et al.<sup>47</sup> and Hansch and Fujita<sup>1</sup> were taken to illustrate the usefulness of the modified Free-Wilson approach to derive significant Hansch equations. No parabolic dependence of log  $A$  on lipophilic character could be found for this group of compounds.

Since the paper of Bruice et al.<sup>47</sup> and Hansch and Fujita's reinvestigation<sup>1</sup> of these data, a number of additional thyronine derivatives have been synthesized and tested,<sup>42,43,46,50-52</sup> therefore the attempt was made to establish quantitative structure-activity correlations for the complete set of compounds. Structures and log  $A$  values are given in Table XII (note that some log  $A$  values are different from those of ref 8 because newer values<sup>43</sup> are used); the parameters used for Hansch analysis and

Table XII. Thyroxine-Like Activity of L-Thyronine Derivatives (Structure I) in the Rat (L-Thyroxine = 100%, assuming  $L = 2 \times DL^{43}$ )

Compd no.	X <sub>1</sub> <sup>a</sup>	X <sub>2</sub> <sup>a</sup>	Y <sub>1</sub>	Y <sub>2</sub>	Mean biol act. in the rat	Log A	Ref
1	H	OH	I	I	1.5	0.18	43
2	H	H	I	I	10	1.00	43
3	H	F	I	I	8	0.90	43
4	H	Cl	I	I	27	1.43	43
5	H	Br	I	I	138	2.14	43
6	H	I	I	I	545	2.74	43
7	H	CH <sub>3</sub>	I	I	126	2.10	43
8	H	C <sub>2</sub> H <sub>5</sub>	I	I	506	2.70	43
9	H	CH(CH <sub>3</sub> ) <sub>2</sub>	I	I	868	2.94	43
10	H	C(CH <sub>3</sub> ) <sub>3</sub>	I	I	120	2.08	43
11	H	CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	I	I	60	1.78	43
12	H	C <sub>6</sub> H <sub>5</sub>	I	I	32	1.51	43
13	F	F	I	I	1.8	0.26	47
14	Cl	Cl	I	I	0.9	-0.05	47
15	Br	Br	I	I	10	1.00	47
16	I	I	I	I	100 <sup>b</sup>	2.00	<i>b</i>
17	F	I	I	I	28	1.45	47
18	H	Br	Br	Br	44	1.64	47
19	H	I	Br	Br	130	2.11	47
20	H	CH(CH <sub>3</sub> ) <sub>2</sub>	Br	Br	450	2.65	50
21	Br	Br	Br	Br	0.51	-0.29	47
22	I	I	Br	Br	12.5	1.10	47
23	H	I	CH <sub>3</sub>	I	40	1.60	52
24	H	I	CH <sub>3</sub>	CH <sub>3</sub>	5	0.70	46
25	H	CH <sub>3</sub>	CH <sub>3</sub>	CH <sub>3</sub>	3	0.48	46
26	H	CH(CH <sub>3</sub> ) <sub>2</sub>	CH <sub>3</sub>	CH <sub>3</sub>	18	1.26	46
27	CH <sub>3</sub>	CH <sub>3</sub>	CH <sub>3</sub>	CH <sub>3</sub>	2	0.30	46
28	I	I	Cl	Cl	0.1	-1.00	47

<sup>a</sup> In case of different substituents X<sub>1</sub> and X<sub>2</sub> the smaller substituent was taken as X<sub>1</sub> and the larger substituent was taken as X<sub>2</sub>. <sup>b</sup> Reference compound.

Table XIII. Parameters Used for Hansch Analysis and Mixed Analysis of Thyronine Derivatives (Structures see Table XII)<sup>a</sup>

Compd no.	$\pi_x^b$	$\sigma_x^c$	$E_{s,X_1}^d$	$E_{s,X_2}^d$	$E_{s,X_2}^{corr e}$	$E_s^f$	[I] <sup>g</sup>	[CH <sub>3</sub> ] <sup>g</sup>
1	-0.67	-0.37	1.24	0.69	0	1.24	2	0
2	0.00	0.00	1.24	1.24	0	1.24	2	0
3	0.14	0.06	1.24	0.78	0	1.24	2	0
4	0.71	0.23	1.24	0.27	0	1.24	2	0
5	0.86	0.23	1.24	0.08	0	1.24	2	0
6	1.12	0.18	1.24	-0.16	0	1.24	2	0
7	0.56	-0.17	1.24	0.00 <sup>j</sup>	0	1.24	2	0
8	1.02	-0.15	1.24	-0.07 <sup>j</sup>	0	1.24	2	0
9	1.53	-0.15	1.24	-0.47 <sup>j</sup>	-0.31	0.93	2	0
10	1.98	-0.20	1.24	-1.54 <sup>j</sup>	-1.38	-0.14	2	0
11	2.00 <sup>h</sup>	-0.12 <sup>i</sup>	1.24	-0.93 <sup>j</sup>	-0.77	0.47	2	0
12	1.96	-0.01	1.24	-2.58	-2.42	-1.18	2	0
13	0.28	0.12	0.78	0.78	0	0.78	2	0
14	1.42	0.46	0.27	0.27	0	0.27	2	0
15	1.72	0.46	0.08	0.08	0	0.08	2	0
16	2.24	0.36	-0.16	-0.16	0	-0.16	2	0
17	1.26	0.24	0.78	-0.16	0	0.78	2	0
18	0.86	0.23	1.24	0.08	0	1.24	0	0
19	1.12	0.18	1.24	-0.16	0	1.24	0	0
20	1.53	-0.15	1.24	-0.47 <sup>j</sup>	-0.31	0.93	0	0
21	1.72	0.46	0.08	0.08	0	0.08	0	0
22	2.24	0.36	-0.16	-0.16	0	-0.16	0	0
23	1.12	0.18	1.24	-0.16	0	1.24	1	1
24	1.12	0.18	1.24	-0.16	0	1.24	0	2
25	0.56	-0.17	1.24	0.00 <sup>j</sup>	0	1.24	0	2
26	1.53	-0.15	1.24	-0.47 <sup>j</sup>	-0.31	0.93	0	2
27	1.12	-0.34	0.00 <sup>j</sup>	0.00 <sup>j</sup>	0	0.00	0	2

<sup>a</sup> Compound 28 was excluded to avoid single point determinations in the mixed analyses. <sup>b</sup> Reference 15. <sup>c</sup>  $\sigma_p$  values, ref 15. <sup>d</sup> Reference 38 unless otherwise indicated. <sup>e</sup> See eq 54. <sup>f</sup> Sum of  $E_{s,X_1}$  and  $E_{s,X_2}^{corr}$ , see eq 57. <sup>g</sup> Free-Wilson parameters, group contributions based on  $\alpha_{Br} = 0.00$ . <sup>h</sup> Estimated value. <sup>i</sup> Reference 53. <sup>j</sup> Reference 54.

mixed analysis are given in Table XIII.

For the choice of substituent parameters some assumptions and approximations have to be made; since  $\pi_{ortho}$  values are known only for a limited number of substituents,  $\pi$  values from the benzene system<sup>15</sup> were used for the calculations (use of the  $\pi_{ortho}$  values from the phenol system<sup>14</sup> for substituents F, Cl, Br and I gave very

similar results in all correlations); as  $\sigma$  values,  $\sigma_{para}$  values were used for the ortho substituents (note that  $\sigma_p = 0.18$  was used for I<sup>15,55</sup>).

Hansch analysis of compounds 1-8 (only trisubstituted compounds; compounds 1-9 or 1-10 or 1-12 gave no equations with statistically significant  $\sigma_x$  terms) gives eq 53. The correlation of observed and calculated log A

Table XIV. Comparison of Observed and Calculated Log A Values

Compd no.	Log A obsd	Log A, calculated from							
		Eq 53	$\Delta$	Eq 55	$\Delta$	Eq 59	$\Delta$	Eq 61	$\Delta$
1	0.18	0.19	0.01	0.19	0.01	0.18	0.00	0.39	0.21
2	1.00	0.83	-0.17	0.86	-0.14	0.80	-0.20	0.86	-0.14
3	0.90	0.99	0.09	1.02	0.12	0.96	0.06	0.98	0.08
4	1.43	1.75	0.32	1.76	0.33	1.69	0.26	1.61	0.18
5	2.14	2.01	-0.13	2.01	-0.13	1.94	-0.20	1.84	-0.30
6	2.74	2.54	-0.20	2.51	-0.23	2.46	-0.28	2.33	-0.41
7	2.10	2.05	-0.05	2.00	-0.10	1.99	-0.11	2.00	-0.10
8	2.70	2.83	0.13	2.75	0.05	2.75	0.05	2.69	-0.01
9	2.94	3.72 <sup>a</sup>	0.78	3.07	0.13	3.09	0.15	3.03	0.09
10	2.08			2.05	-0.03	2.10	0.02	2.22	0.14
11	1.78			3.03 <sup>b</sup>	1.25	c		d	
12	1.51			0.00 <sup>b</sup>	-1.51	c		d	
13	0.26					0.32	0.06	0.42	0.16
14	-0.05					0.92 <sup>c</sup>	0.97	0.91 <sup>d</sup>	0.96
15	1.00					1.11	0.11	1.09	0.09
16	2.00					1.72	-0.28	1.71	-0.29
17	1.45					1.83	0.38	1.77	0.32
18	1.64							1.49	-0.15
19	2.11							1.98	-0.13
20	2.65							2.68	0.03
21	-0.29							0.74 <sup>d</sup>	1.03
22	1.10							1.36	0.26
23	1.60							1.59	-0.01
24	0.70							0.85	0.15
25	0.48							0.52	0.04
26	1.26							1.55	0.29
27	0.30							-0.18	-0.48

<sup>a</sup> Not included in calculation of eq 53. <sup>b</sup> Not included in calculation of eq 55. <sup>c</sup> Not included in calculation of eq 59. <sup>d</sup> Not included in calculation of eq 61.

$$\log A = 1.746 (\pm 0.421) \pi_x - 1.410 (\pm 1.176) \sigma_x + 0.834 \quad (53)$$

$n = 8; r = 0.982; s = 0.207; F = 66.42; p < 0.001$

values is very good (see Table XIV); eq 53 is the quantitative expression of the statements made by Jorgensen<sup>43</sup> that biological activity increases with lipophilic character and that for equal lipophilic character the electron-releasing alkyl substituents lead to more activity than the electron-attracting halogen substituents (the anomalous position of the fluorine is caused by its smaller  $\sigma$  value).

If eq 53 is used to calculate the log A value of compound 9, the value obtained is much too high (log A calcd = 3.72; see Table XIV); for compounds 10-12 even higher log A values are calculated from eq 53. Therefore, an additional term has to be included in eq 53 to account for the lower activity of compounds with large, lipophilic X<sub>2</sub> substituents.

Two different factors may be responsible for this lower activity: a parabolic dependence of log A values on lipophilic character and/or a steric hindrance for substituents larger than iodine. Although the use of an additional  $\pi^2$  term gave significant correlations for compounds 1-12 ( $n = 12; r = 0.914; s = 0.394; F = 13.61; p < 0.005$ ) as well as for compounds 1-17 ( $n = 17; r = 0.753; s = 0.658; F = 5.69; p < 0.05$ ), a  $\pi^2$  term was without any significance when compounds 18-27 were included in the calculations (compare ref 8).

The use of  $E_{s,X_2}$  as an additional parameter gave no correlations with significant  $E_{s,X_2}$  terms; this is not surprising because eq 53 indicated no dependence of biological activity on the size of the X<sub>2</sub> substituent for groups similar in size or smaller than iodine. Only substituents larger than iodine cause lower than predicted activities. Therefore the hypothesis was made that the thyroxine receptor has a "pocket" for the X<sub>2</sub> substituent, great enough for the iodine atom, which is the X<sub>2</sub> substituent of the natural hormone, and for substituents with equal

or smaller size. To get a quantitative measure for this hypothesis that only substituents larger than iodine cause steric hindrance, modified  $E_s$  values were defined.

$$E_{s,X_2}^{\text{corr}} = E_{s,X_2} - E_{s,I} = E_{s,X_2} + 0.16 \quad (54)$$

(Only negative values of  $E_{s,X_2}^{\text{corr}}$  are considered; positive values indicate that there is no steric hindrance; therefore all positive values were put to zero.)

With these modified  $E_s$  values (accounting only for the steric hindrance of X<sub>2</sub> substituents larger than iodine) eq 55 is obtained for compounds 1-10.

$$\log A = 1.673 (\pm 0.324) \pi_x - 1.242 (\pm 0.969) \sigma_x + 1.714 (\pm 0.600) E_{s,X_2}^{\text{corr}} + 0.856 \quad (55)$$

$n = 10; r = 0.984; s = 0.201; F = 59.53; p < 0.001$

Again the correlation of observed and calculated log A values is very good (see Table XIV). Greater deviations are obtained if log A values for compounds 11 and 12 are calculated from eq 55, but these deviations may be caused by the uncertainties of  $E_s$  values for the isobutyl group (compound 11) and the phenyl group (compound 12).

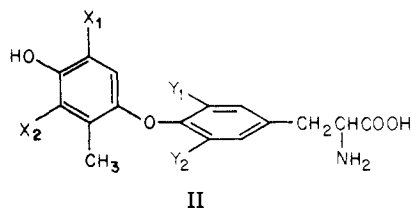
The isobutyl group: if a "pocket" for the X<sub>2</sub> substituent is assumed, the steric hindrance for an isobutyl group may be of the same order or even greater than for a *tert*-butyl group; for normal chemical reactions the steric hindrance of an isobutyl group is smaller than that of a *tert*-butyl group due to the greater flexibility of the isobutyl group.

The phenyl group: Kutter and Hansch<sup>38</sup> calculated two different  $E_s$  values for the phenyl group:  $E_s = -2.58$  (function coplanar to reaction center) and  $E_s = 0.23$  (function perpendicular to reaction center). Although the actual  $E_s$  value will be nearer the value of -2.58 there remains a high degree of uncertainty about the real  $E_s$  value as long as the structure of the receptor is unknown.

For these reasons compounds 11 and 12 were excluded from the following calculations (it must be noted that

Greenberg et al.<sup>42</sup> discussed the relationships between the bulk of the X<sub>2</sub> substituent and the biological activity for several alkyl-substituted thyronines).

Tetrasubstituted thyronines (e.g., compounds 13–17, 21, 22) are less active than the corresponding trisubstituted thyronines. While Hansch and Fujita<sup>1</sup> tried to correlate this fact with a parabolic dependence of log *A* on lipophilic character, Jorgensen<sup>41,43</sup> assumed a steric hindrance for the second X substituent (X<sub>1</sub> in structures I and II) and gave an excellent evidence for this assumption; he investigated 2'-methylthyronines (structure II), which are fixed in the presented conformation due to the steric



interactions between the 2'-methyl group and the aromatic ring bearing the Y substituents. While a 2',3'-dimethyl-3,5-diiodo-DL-thyronine (X<sub>1</sub> = H; X<sub>2</sub> = CH<sub>3</sub>) with a distally oriented 3'-methyl group shows 50% the activity of L-thyroxine, the isomeric 2',5'-dimethyl-3,5-diiodo-DL-thyronine (X<sub>1</sub> = CH<sub>3</sub>; X<sub>2</sub> = H) with a proximally oriented 5'-methyl group shows only 1% the activity of L-thyroxine. This steric hindrance for the second X substituent is completely confirmed by eq 52 (see ref 8).

Therefore  $\pi_x$ ,  $\sigma_x$ , and  $E_{s,X_1}$  were used to derive eq 56 for

$$\log A = 1.908 (\pm 0.517) \pi_x - 2.151 (\pm 1.517) \sigma_x + 1.871 (\pm 0.700) E_{s,X_1} - 1.598 \quad (56)$$

$n = 13; r = 0.946; s = 0.347; F = 25.59; p < 0.001$

compounds 1–8 and 13–17. Because the coefficients of the  $E_s$  terms are nearly identical in eq 55 and 56 both  $E_s$  values were combined according to eq 57 (the numerical equiv-

$$E_s' = E_{s,X_1} + E_{s,X_2}^{\text{corr}} \quad (57)$$

alence of the coefficients may be taken as evidence for the biological significance of the  $E_{s,X_2}^{\text{corr}}$  term in eq 55).  $E_s'$  can be interpreted as a quantitative measure for the steric hindrance of X<sub>1</sub> substituents and of X<sub>2</sub> substituents larger than iodine. Using  $\pi_x$ ,  $\sigma_x$ , and  $E_s'$  eq 58 was derived for

$$\log A = 1.787 (\pm 0.388) \pi_x - 1.787 (\pm 0.863) \sigma_x + 1.858 (\pm 0.552) E_s' - 1.551 \quad (58)$$

$n = 15; r = 0.953; s = 0.329; F = 36.19; p < 0.001$

compounds 1–10 and 13–17. If compound 14 is excluded, a similar equation with a somewhat higher correlation coefficient is obtained for compounds 1–10, 13, and 15–17.

$$\log A = 1.704 (\pm 0.271) \pi_x - 1.398 (\pm 0.636) \sigma_x + 1.709 (\pm 0.389) E_s' - 1.319 \quad (59)$$

$n = 14; r = 0.975; s = 0.223; F = 65.55; p < 0.001$

A comparison of observed log *A* values and log *A* values calculated from eq 59 is given in Table XIV.

To establish quantitative structure–activity correlations for all compounds of Table XII the mixed approach (eq 8) with Hansch parameters  $\pi_x$ ,  $\sigma_x$ , and  $E_s'$  for substituents X<sub>1</sub> and X<sub>2</sub> and Free–Wilson parameters [I] and [CH<sub>3</sub>] (group contributions of I and CH<sub>3</sub>, based on  $\alpha_{Br} = 0.00$ ) for substituents Y<sub>1</sub> and Y<sub>2</sub> was used; compounds 11 and

Table XV. Dependence of Biological Activity on Substituents Y

Y <sub>1</sub> , Y <sub>2</sub>	Group contributions, based on $\alpha_{Br} = 0.00^a$	Rel biol act. in % (based on Y <sub>1</sub> = Y <sub>2</sub> = I, 100%)
I	0.176	100
Br	0.00	44
CH <sub>3</sub>	-0.563	3
Cl	ca. -1.0 <sup>b</sup>	0.4

<sup>a</sup> From eq 61. <sup>b</sup> Estimated from compounds 22 and 28.

12 were excluded for the reasons discussed above; compound 28 was excluded to avoid a single point determination.

$$\log A = 1.699 (\pm 0.336) \pi_x - 2.059 (\pm 0.702) \sigma_x + 1.713 (\pm 0.388) E_s' + 0.234 (\pm 0.201) [I] - 0.532 (\pm 0.261) [CH_3] - 1.792 \quad (60)$$

$(t = 10.59) \quad (t = 6.14)$   
 $(t = 9.25) \quad (t = 2.44)$   
 $(t = 4.26)$

$n = 25; r = 0.943; s = 0.349; F = 30.37; p < 0.001$

If compounds 14 and 21 are excluded, eq 61 with similar

$$\log A = 1.569 (\pm 0.251) \pi_x - 1.582 (\pm 0.555) \sigma_x + 1.493 (\pm 0.299) E_s' + 0.176 (\pm 0.159) [I] - 0.563 (\pm 0.195) [CH_3] - 1.348 \quad (61)$$

$(t = 13.21) \quad (t = 6.02)$   
 $(t = 10.54) \quad (t = 2.34)$   
 $(t = 6.10)$

$n = 23; r = 0.965; s = 0.250; F = 46.53; p < 0.001$

coefficients but a somewhat higher correlation coefficient is obtained for compounds 1–10, 13, 15–20, and 22–27.

All parameters of eq 60 and 61 are statistically significant (the *t* values are given in parentheses under each term); a certain degree of collinearity exists between  $\pi_x$  and  $E_s'$  ( $r^2 = 0.471$ ) but no way could be found to circumvent this problem. The correlation between observed and calculated log *A* values (see Table XIV) is good if one considers that most biological activities are mean values from different biological test models and different investigators, summed up over a period of more than 30 years of research in this field.

From eq 60 and 61 the following conclusions can be drawn. (a) Biological activity increases with lipophilic character; no parabolic dependence of log *A* on lipophilic character can be seen. (b) Electron-releasing properties of a substituent increase the biological activity, electron-attracting properties decrease the biological activity. (c) No steric hindrance can be seen for 3'-substituents (X<sub>2</sub> substituents) as far as bromine, iodine, methyl, ethyl, or smaller substituents are concerned; 3'-substituents larger than iodine and all 5'-substituents (X<sub>1</sub> substituents) lead to a decrease of biological activity due to steric hindrance; this corresponds to a high specificity of the thyroid hormone receptor for the natural hormone 3',3,5-triiodothyronine. Only for X<sub>2</sub> = isopropyl the greater + $\pi$  and - $\sigma$  effects surpass the deactivating effect due to steric hindrance; therefore the 3'-isopropyl compound is even more active than the natural hormone; for X<sub>2</sub> = isobutyl, *tert*-butyl, or larger substituents the steric hindrance surpasses all other effects. (d) Compounds with Y<sub>1</sub> = Y<sub>2</sub> = I are the most active analogues; activity decreases in the order I > Br > CH<sub>3</sub> > Cl (see Table XV).

Compounds with Y = H, F, *i*-Pr, *s*-Bu, CN, COOH, NO<sub>2</sub>, NH<sub>2</sub>, SC<sub>2</sub>H<sub>5</sub>, or SC<sub>6</sub>H<sub>5</sub> are inactive;<sup>43,56</sup> that means that large lipophilic substituents like bromine or iodine lead to biologically active compounds, while smaller substituents or substituents larger than iodine or polar substituents cause a significant reduction or loss of activity. While the importance of bulky substituents like bromine or iodine for a favorable conformation is well understood from MO calculations,<sup>45,57</sup> the most reasonable explanation for the inactivity of compounds with Y = *i*-Pr or *s*-Bu seems to be a steric hindrance which prevents binding to a receptor molecule.<sup>45</sup> This specific structure-activity relationship would be best described by a  $+π_y/-E_{s,Y}/+E_{s,Y}^2$  correlation but not enough appropriate data are available to derive such a correlation.

### Discussion

The two most frequently used models in quantitative structure-activity relationships, Hansch's linear multiple regression model and Free-Wilson's additive model, have been combined in eq 7, 8, and 9 to a mixed approach, based on the equivalence<sup>8</sup> of both models.

The preceding examples have been discussed in detail to demonstrate the use and the advantages of the mixed approach (a) to check the fitting of a particular Hansch equation, (b) to derive significant Hansch equations, (c) to detect parabolic dependence of log 1/*C* values on lipophilic character also in Free-Wilson analysis, (d) to apply Free-Wilson analysis in such cases, (e) to describe structure-activity relationships for compounds with great structural variation in one position and little structural variation in another position of the molecule with the smallest number of variables possible, and (f) to describe structure-activity relationships in cases where strong collinearity of structural parameters renders the derivation of meaningful Hansch correlations impossible.

In appropriate cases the mixed approach can be used as a tool to derive statistically significant and biologically meaningful correlations; some care should be taken in the right selection of Free-Wilson parameters in order not to generate meaningless mixtures of structural parameters and dummy variables (e.g., for different substituents in only one position of a molecule); a minimum number of compounds per variable should be present to avoid chance correlations.<sup>39</sup> In every case the final equation should be checked with the criteria given by Unger and Hansch<sup>58</sup> (no *t* test or sequential *F* test should be applied to the Free-Wilson parameters of a mixed equation because also a group contribution of zero is meaningful despite its *t* value).

If these points are taken into account, the mixed approach widens the scope and diminishes the limitations of the linear free energy related Hansch model and the mathematical Free-Wilson model; the mixed approach "incorporates the best aspects of the two quantitative approaches" as Hansch<sup>22</sup> has said in his definition of an equation including structural parameters and Free-Wilson-like dummy variables.

**Acknowledgment.** I am grateful to O. H. Kehrhan for helpful discussions and for computational assistance.

### References and Notes

- (1) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
- (2) C. Hansch, *Acc. Chem. Res.*, **2**, 232 (1969).
- (3) C. Hansch in "Drug Design", Vol. I, E. J. Ariens, Ed., Academic Press, New York, N.Y., 1971, p 271.
- (4) A. Cammarata and S. J. Yau, *J. Med. Chem.*, **13**, 93 (1970).
- (5) T. Fujita and T. Ban, *J. Med. Chem.*, **14**, 148 (1971).
- (6) S. M. Free, Jr., and J. W. Wilson, *J. Med. Chem.*, **7**, 395 (1964).

- (7) A. Cammarata, *J. Med. Chem.*, **15**, 573 (1972).
- (8) Paper 1: H. Kubinyi and O. H. Kehrhan, *J. Med. Chem.*, preceding paper in this issue.
- (9) C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1973).
- (10) J. A. Singer and W. P. Purcell, *J. Med. Chem.*, **10**, 1000 (1967).
- (11) K. Bocek, J. Kopecký, M. Krivucová, and D. Vlachová, *Experientia*, **20**, 667 (1964).
- (12) J. Kopecký, K. Bocek, and D. Vlachová, *Nature (London)*, **207**, 981 (1965).
- (13) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71**, 525 (1971).
- (14) T. Fujita, J. Isawa, and C. Hansch, *J. Am. Chem. Soc.*, **86**, 5175 (1964).
- (15) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).
- (16) C. Hansch, A. Leo, and D. Nikaitani, *J. Org. Chem.*, **37**, 3090 (1972).
- (17) A. Canas-Rodriguez and M. S. Tute, *Adv. Chem. Ser.*, **No. 114**, 41 (1972).
- (18) P. N. Craig, *Adv. Chem. Ser.*, **No. 114**, 115 (1972).
- (19) P. N. Craig and C. Hansch, *J. Med. Chem.*, **16**, 661 (1973).
- (20) T. Ban and T. Fujita, *J. Med. Chem.*, **12**, 353 (1969).
- (21) P. N. Craig, *J. Med. Chem.*, **15**, 144 (1972).
- (22) C. Hansch and M. Yoshimoto, *J. Med. Chem.*, **17**, 1160 (1974).
- (23) C. Hansch and E. J. Lien, *Biochem. Pharmacol.*, **17**, 709 (1968).
- (24) Y. C. Martin, *J. Med. Chem.*, **13**, 145 (1970).
- (25) C. Hansch, *J. Org. Chem.*, **35**, 620 (1970).
- (26) J. W. Mc Farland, *Prog. Drug Res.*, **15**, 123 (1971).
- (27) Y. C. Martin and K. R. Lynn, *J. Med. Chem.*, **14**, 1162 (1971).
- (28) C. Hansch and W. R. Glave, *J. Med. Chem.*, **15**, 112 (1972).
- (29) Y. C. Martin, P. H. Jones, T. J. Perun, W. E. Grundy, S. Bell, R. R. Bower, and N. L. Shipkowitz, *J. Med. Chem.*, **15**, 635 (1972).
- (30) T. Fujita, *J. Med. Chem.*, **16**, 923 (1973).
- (31) Y. C. Martin, J. B. Holland, C. H. Jarboe, and N. Plotnikoff, *J. Med. Chem.*, **17**, 409 (1974).
- (32) C. Silipo and C. Hansch, *Mol. Pharmacol.*, **10**, 954 (1974).
- (33) J. K. Seydel, H. Ahrens, and W. Losert, *J. Med. Chem.*, **18**, 234 (1975).
- (34) F. R. Quinn, J. S. Driscoll, and C. Hansch, *J. Med. Chem.*, **18**, 332 (1975).
- (35) R. Franke and P. Oehme, *Pharmazie*, **28**, 489 (1973).
- (36) E. J. Lien, C. Hansch, and S. M. Anderson, *J. Med. Chem.*, **11**, 430 (1968).
- (37) C. Hansch and E. J. Lien, *J. Med. Chem.*, **14**, 653 (1971).
- (38) E. Kutter and C. Hansch, *J. Med. Chem.*, **12**, 647 (1969).
- (39) J. G. Topliss and R. J. Costello, *J. Med. Chem.*, **15**, 1066 (1972).
- (40) M. S. Tute, *Adv. Drug. Res.*, **6**, 1 (1971).
- (41) E. C. Jorgensen, P. A. Lehman, C. Greenberg, and N. Zenker, *J. Biol. Chem.*, **237**, 3832 (1962).
- (42) C. M. Greenberg, B. Blank, F. R. Pfeiffer, and J. F. Pauls, *Am. J. Physiol.*, **205**, 821 (1963).
- (43) E. C. Jorgensen in "Medicinal Chemistry", Vol. II, A. Burger, Ed., Wiley-Interscience, New York, N.Y., 1970, p 838.
- (44) P. A. Lehmann, *J. Med. Chem.*, **15**, 404 (1972).
- (45) P. A. Kollman, W. J. Murray, M. E. Nuss, E. C. Jorgensen, and S. Rothenberg, *J. Am. Chem. Soc.*, **95**, 8518 (1973).
- (46) E. C. Jorgensen, W. J. Murray, and P. Block, Jr., *J. Med. Chem.*, **17**, 434 (1974).
- (47) T. C. Bruice, N. Kharasch, and R. J. Winzler, *Arch. Biochem. Biophys.*, **62**, 305 (1956).
- (48) G. Cilento and M. Berenholz, *Biochem. Biophys. Acta*, **94**, 271 (1965).
- (49) C. Hansch, A. R. Steward, J. Isawa, and E. W. Deutsch, *Mol. Pharmacol.*, **1**, 205 (1965).
- (50) E. C. Jorgensen and J. R. Nulu, *J. Pharm. Sci.*, **58**, 1139 (1969).
- (51) E. C. Jorgensen and P. Block, Jr., *J. Med. Chem.*, **16**, 306 (1973).
- (52) E. C. Jorgensen and R. A. Wiley, *J. Med. Pharm. Chem.*, **5**, 1307 (1962).

- (53) H. H. Jaffé, *Chem. Rev.*, **53**, 191 (1953).  
 (54) R. W. Taft, Jr., in "Steric Effects in Organic Chemistry", M. S. Newman, Ed., Wiley, New York, N.Y., 1956, p 556.  
 (55) D. H. McDaniel and H. C. Brown, *J. Org. Chem.*, **23**, 420 (1958).  
 (56) E. C. Jorgensen, R. O. Muhlhauser, and R. A. Wiley, *J. Med. Chem.*, **12**, 689 (1969).  
 (57) L. B. Kier and J. R. Hoyland, *J. Med. Chem.*, **13**, 1182 (1970).  
 (58) S. H. Unger and C. Hansch, *J. Med. Chem.*, **16**, 745 (1973).

## Quantitative Structure-Activity Studies on Monoamine Oxidase Inhibitors

Carl Lynn Johnson

Department of Pharmacology, Mount Sinai School of Medicine, City University of New York, New York, New York 10029.  
 Received June 5, 1975

Quantitative structure-activity studies were carried out on a series of *N*-isopropylaryl hydrazides which inhibits monoamine oxidase (MAO). The inhibitory potencies of these compounds on MAO were found to correlate with the electron-withdrawing capacity of the aryl ring substituents as estimated by both empirical Hammett  $\sigma$  constants and electronic indices from molecular orbital calculations. Based on these correlations and previously published data on other classes of MAO inhibitors, a general model for the inhibitor pharmacophore is proposed: potent MAO inhibitors should contain an electron-rich functional group in the plane of and approximately 5.3 Å from the center of an aromatic ring; electron-withdrawing groups on the aromatic ring or replacing the phenyl ring with certain types of heterocyclic rings will tend to increase the potency.

The importance of monoamine oxidase (MAO) in controlling the levels of neurotransmitters has prompted interest in learning the mechanism of substrate oxidation and the nature of the active site as well as in understanding the molecular mechanisms of MAO inhibition by a wide variety of chemical compounds, all of which exhibit a pharmacological profile including potent antidepressant activity. Some of the known MAO inhibitors<sup>1,2</sup> include arylalkylhydrazines, aryl hydrazides, arylpropargylamines, arylcyclopropylamines, aryloxycyclopropylamines, *N*-cyclopropylaryloxyethylamines,  $\beta$ -carbolines, and  $\alpha$ -methylated arylalkylamines. A common structural feature of both substrates and these various classes of inhibitors is an amino or imino group, which is assumed to play an essential role in complex formation at the active site. The aryl portion of the MAO inhibitors is not an absolute requirement for inhibition of the enzyme, but substituents on the aryl ring or modification of the type of ring markedly affect potency.

One approach to understanding the role of the aryl portion of these compounds in MAO inhibition and to gain insight into the relationships between chemical structure and pharmacological activity is to carry out quantitative structure-activity relationships (QSAR)<sup>3</sup> on a series of related molecules. This approach requires, in addition to the estimate of the biological activity of the molecules, a set of parameters—either empirically or theoretically derived—which describes in numerical terms the structural characteristics of the compounds. Several QSAR studies have dealt with both substrates<sup>4</sup> and inhibitors<sup>5-7</sup> of MAO. The study of Fujita<sup>7</sup> is especially extensive, comparing the QSAR results for many classes of inhibitors.

One group of MAO inhibitors not previously considered in a QSAR study is the aryl hydrazides. In the present communication we report on the QSAR for a series of hydrazides and compare the results with those for other classes of MAO inhibitors in an attempt to arrive at a general picture of the inhibitory pharmacophore. A preliminary report of some of these results has appeared.<sup>8</sup>

**Methods.** Molecular orbital (MO) calculations were carried out by the CNDO/2 method<sup>9</sup> with the CNINDO computer program.<sup>10</sup> Coordinates were generated with the MBLD program<sup>11</sup> from crystallographic data and/or standard bond angles and bond lengths.<sup>12</sup> Calculations were performed on a CDC 6600 computer. Multiple re-

gression analysis was carried out with a stepwise regression program adapted for the PROPHET<sup>13</sup> time-sharing system. Hammett  $\sigma$  constants were taken from McDaniel and Brown<sup>14</sup> and octanol-H<sub>2</sub>O partition coefficients were taken from the compilations of Hansch and co-workers.<sup>15,16</sup>

### Results and Discussion

**Empirical Structure-Activity Relationships.** The *in vivo* potencies of a series of 23 aryl hydrazides of the general structure ArCONHNHCH(CH<sub>3</sub>)<sub>2</sub> have been published<sup>17</sup> and are listed in Table I. The potencies are reported as the Marsilid Index (M.I.) taking iproniazid (4-pyridyl derivative) as the standard (M.I. = 100). The Marsilid Index, designed to measure the relative degree of inhibition of brain MAO, is defined as the ratio of the increase in serotonin in rat brain produced by a substance (in amount equimolar to 100 mg of iproniazid/kg) to the increase in serotonin produced by 100 mg of iproniazid/kg.

The electronic substituent constant,  $\sigma$ , and the hydrophobic or lipid solubility constant,  $\log P$ , are listed in Table I. Since Hammett  $\sigma$  constants are available only for the phenyl derivatives, the empirical QSAR studies were restricted to these ten compounds. Multiple regression analysis led to eq 1-7. The numbers in parentheses in the regression equations represent the standard errors of the regression coefficients;  $n$  is the number of compounds;  $r$  is the multiple correlation coefficient;  $s$  is the standard deviation of the regression; and  $F$  is the  $F$  test for the significance of the regression. A significant correlation ( $p < 0.01$ ) of the Marsilid Index and  $\sigma$  values was obtained (eq 1). The correlation with  $\log P$  was much poorer (eq

$$\text{M.I.} = 88.25 (\pm 23.09) \sigma + 83.37 \quad (1)$$

$$n = 10; r = 0.804; s = 31.4; F = 14.6$$

2). Including a  $\log P$  term with  $\sigma$  did not improve the

$$\text{M.I.} = 32.80 (\pm 13.74) \log P + 0.64 \quad (2)$$

$$n = 10; r = 0.646; s = 30.3; F = 5.7$$

correlation (eq 3). The major deviations from the above

$$\text{M.I.} = 70.51 (\pm 24.41) \sigma + 17.13 (\pm 11.30) \log P + 44.33 \quad (3)$$

$$n = 10; r = 0.856; s = 29.1; F = 9.6$$