Chapter **33.** Physicochemical Parameters in Drug Design Corwin Hansch, Pomona College, Claremont, California

Introduction - Interest in quantitative correlation of chemical structure and biological activity is developing at a much more rapid pace than one might deduce from the number of papers which have appeared since our last review.' From informal talks it is apparent that a variety of laboratories have formulated computer programs and are actively pursuing such studies. The two general approaches that are receiving attention are the empirical Free-Wilson $^2$  and Kopeck $\acute{\textbf{y}}$ -Bo $\texttt{\textcolor{red}{\check{c}}}$ ek $^5$  methods and the semi-empirical method in which changes in biological activity of members of a set of congeners are correlated with known physicochemical constants.<sup>1</sup> We are unquestionably entering a new era where the role of the computer in drug design is becoming indispensible. As pharmacological testing becomes more precisely defined in numerical values, the drug designer faces the sharper challenge of explaining quantitative differences in physical chemical terms. Sorting out roles for the large number of variables can only be accomplished through computerized regression analysis.

*De Novo* Constants - Purcell and his colleagues have continued their development of the Free-Wilson empirical method in which the biological activity is expressed as a function of the activity contributions associated with segments of the molecule (substitutent groups and the parent portion). In one study3 they have obtained *de novo* substituent constants for the contribution of hydrogen, methoxy, and various alkyl groups to the hypoglycemic activity of piperidinesulfamylsemicarbazides. From the observed biological activity of 12 derivatives they have predicted activity for 12 other derivatives. In a second study they have derived substituent constants for group contributions to the antitumor activity of acetylenic carbamates. *de ~OVO* constants for the correlation of toxicities to mice of ortho-disubstituted benzenes. Bocek and Kopecky have reported<sup>5</sup> an example of the use of

Semi-Empirical Method - More effort is being directed toward using known physicochemical parameters to rationalize the change in biological activity resulting from the modification of a parent molecule. In this approach, attempts are being made to relate **ABR** (change in biological response) to changes in hydrophobic, electronic, and steric effects of substituents. The very great importance of hydrophobic interactions in biochemical systems is receiving increased attention.<sup>6</sup>

Partition Coefficients - In using an aqueous phase and a simple organic solvent to serve as a reference system for approximating hydrophobic interactions in biochemical systems, Collander<sup>46</sup> has pointed out that we are assuming: log  $P_1 = a \log P_2 + b$ . In this relationship  $P_1$  is the partition coefficient in one system (fatty and aqueous biophases) and  $P_2$  is the partition coefficient in a second system (e.g., octanol-water). While Collander did provide some evidence for this linear relationship, much

Chap. 33 Drug Design Hansch 349 better evidence is available from the earlier work of H. W. Smith. Equa-

 $log P_{CHC1_2} = 1.064 log P_{xylene} + 0.393$  n = 30 r = 0.979 (1)

tion 1, based on Smith's results, shows<sup>7</sup> a good linear relationship between the partition coefficients of a variety of acids in the two systems CHCl<sub>3</sub>-water and xylene-water. In eq 1, n is the number of compounds used in the regression and r is the correlation coefficient. Other good examples validating Collander's relation come from the work of Kakemi, Arita, Hori and Konishi. $\delta$  These workers also pointed out that for barbiturates the value of P is independent of the concentration of barbiturate partitioned. This seems to be generally true of more or less neutral molecules.<sup>27</sup> Of more interest is the type of correlation possible between a simple solvent system and biopolymers. That log P or  $\pi$  from the octanolwater system is a good parameter for such correlations can be seen from the following equations. In eq 2 and **3,** K are binding constants of pre-

Binding of ROH by Ribonuclease7



cise definition while eq 4 and 5 are not *so* sharply defined. In eq 5, B refers to the % bound and F to the % free. The work of Bird and Marshall<sup>10</sup> embodied in eq 5, is most significant because of the large number of compounds studied and the great variation in their structure. The binding of sulfanilamides to plasma protein has also been quantitatively correlated using substituent constants **.I1** Kakemi, Arita, Hori and Konishi have shown that the absorption of barbiturates by the rat stomach is linearly related to their partition coefficients. $^{12}$  That electronic effects of substituents can outweigh lipophilic effects in certain instances is apparent from a study of the uptake of S-benzoylthiamines by human erythrocytes.<sup>13</sup> In this study a good correlation was found between  $\sigma$  and the adsorption of a given derivative. Actually, in this case **u** may correlate the degree of ionization which affects the partition coefficient.

Enzyme Studies - Log P, along with the Hammett  $\sigma$  constant, has found use in the correlation of enzymic reactions.<sup>1</sup> The study of Blomquist<sup>14</sup> on the electronic effects of substituents in alcohol dehydrogenase reactions illustrates the point. Blomquist noted that both electronic and hydrophobic effects appear to be involved in the enzymic reduction of benzaldehydes. We have placed his results in mathematical context in eq 6 and 7 where K =



 $[1/\phi_2]$  derivative/ $1/\phi_2$  benzaldehyde)100]. Equation 7 is a highly significant improvement over eq 6. An F test indicates the  $\pi$  term to be significant at  $\star$ .995. The  $\pi$  values are from the benzoic acid system.<sup>15</sup> The coefficients in eq 7 are similar to those found by McMahon for the enzymic reduction of aromatic ketones to alcohols.1 Another enzymic example is embodied in eq 8, formulated from the results of Bender *et aZ.16* for the cycloamylose-catalyzed hydrolysis of phenyl acetates. The five data points are for the meta isomers only. Equation 8 is not highly significant statistically because of the few points used in the regression; how-

$$
\log K = 0.506\pi + 1.202\sigma - 1.432 \qquad n = 5 \qquad r = 0.913 \quad (8)
$$

ever, it does point to new directions for research. Wildnauer and Canady have used a variety of parameters to correlate enzyme-inhibitor complexes of a-chymotrypsin **.45** A considerably more complex structure-activity relationship, in terms of  $\pi$  and  $\sigma$ , was found for the inhibition of malate dehydrogenase by phenols. $^{17}$ 

Nonspecific Inhibition - The use of log P in the correlation of the structure-activity relationship in a wide variety of narcotics<sup>9</sup> has shown a linear relation between log 1/C and log P for 16 different systems with slopes near 1. The similar equations point to a common mechanism of action, possibly the inhibition of electron transport in oxidative metabolism. The use of log P for relating the effect of different sets of congeners acting on different biological systems is indicated in eq 9 and 10.

50% Inhibition *Arbacia* Egg Cell Division by Barbiturates9

 $log 1/C = 0.801 log P + 1.076$  n = 19 r = 0.960 (9)

Inhibition *Avena* Cell Elongation by Phenoxyacetic Acids<sup>18</sup>

$$
log 1/C = 0.778 log P + 1.971
$$
  $n = 22$   $r = 0.928 (10)$ 

The slopes of the above equations are the same, indicating the same dependence of inhibition on the hydrophobic character of the drug. The intercepts are different, indicating greater sensitivity of the *Avena* test. These are arbitrary standards and it would be interesting to place both on the same basis to compare the intrinsic activity of the two classes of inhibitors. An equation similar to 10 was also found for the phenylacetic acids. $^{18}$  While both sets of acids are toxic at high concentrations, at much lower concentrations they promote cell elongation.

An interesting example illustrating how conformational changes in an enzyme can be quantitatively correlated using  $\pi$  constants comes from the work of Ichikawa and Yamano.<sup>19</sup> Equation 11 correlates the molar concentration of phenol causing 50% conversion of cytochrome P450 to P420.

$$
log \ 1/C = 0.631\pi + 1.194
$$
  $n = 11$   $r = 0.986$  (11)

The study was made using liver microsomes and 14 different phenols. Equation 11 differs from the one formulated by Ichikawa and Yamano in that the three most lipophilic phenols (2,4,6,-triCl; 2,3,4,6-tetra-C1; penta-C1)

were omitted since a plot of the data showed a departure from linearity at  $\pi$   $\cong$  2.3. Using a  $\pi$ <sup>2</sup> term, all points are well fit. An equation similar to 11 was also found for aniline derivatives.<sup>19</sup>

Alkylating Agents - Hansch and Lien<sup>20</sup> have shown that adrenergic blocking by 8-halophenylethylamines (I) studied by Graham can be correlated by two types of equations.



Equation 12 correlates the effect of ring substituents (X) in compounds of type I and eq 13 correlates the effects of changes in R. In eq 12, where

$$
log 1/C = 1.221\pi - 1.587\sigma + 7.888
$$
  $n = 22$   $r = 0.918$  (12)

$$
\log 1/C = 1.113E_{\rm g}^{\rm c} + 3.566\sigma^* - 4.432n_{\rm H} + 11.911 \quad n = 10 \quad r = 0.986 \tag{13}
$$

the haloamines were studied *US.* adrenaline in rats, the hydrophobic character of the substituent was most important. In eq 13, where the benzene ring was held constant and variations in the R groups attached to the nitrogen of the side chain were made, hydrophobic binding did not appear to play a significant role. In eq 13,  $n_H$  represents the number of hydrogens attached to the protonated nitrogen and E<sub>S</sub> is Hancock's corrected steric parameter.  $E_S^C$  often appears to give better results<sup>7</sup> than  $E_S$ . The rationale for using n<sub>H</sub> was that eq 14 could be used to correlate the base strengths of primary, secondary, and tertiary amines.<sup>20</sup>

$$
pK_a = 3.140\sigma^* + 1.816n_{\mu} + 7.817
$$
   
 
$$
n = 92 \quad r = 0.985 \ (14)
$$

Ishida<sup>21,22</sup> and coworkers have studied the ovicidal activity of a set of congeners of bromoethylthiobenzenes (111). Equation 15 was derived to rationalize the toxicities. They postulated that the onium compound (IV)

$$
log 1/C = -2.24\pi^2 + 1.74\pi - 1.44\sigma + 4.34
$$
  $n = 8$   $r = 0.967$  (15)

**was** an important intermediate and that this intermediate might then act biochemically as an alkylating agent. In an *in vitro* study of the alkylation of **4-(p-nitrobenzyl)-pyridine** with compounds of type 111, they found a rho constant of -1.85 to *-2.04.* Rho for hydrolysis was similar, -1.71 to -1.98. These negative values of rho are reasonably close to  $-1.44$  of eq 15 considering the different "solvent" environments. The val-<br>ue of rho of eq 15 is close to  $-1.59$  of eq 12. The haloalkylamines are also assumed to act as alkylating agents via II. While one would not exue of rho of eq 15 is close to  $-1.59$  of eq 12. The haloalkylamines are

pect the same value for rho in these two systems, similar values would be expected.

Hill Reaction - Equation 16 correlates<sup>23</sup> the inhibiting action of N,N-dialkylphenylureas on the Hill reaction. Equation 16 is similar to that

 $pI_{50} = 1.78\pi - 2.44\sigma + 4.59$   $n = 8$   $r = 0.87$  (16)

found for other Hill reaction inhibitors.<sup>1</sup>

Nonlinear Dependence of Activity on  $Log P$  - Evidence<sup>24,25</sup> continues to mount that while the dependence of biological activity on log P is not linear, over large ranges of P it can be treated in a rational manner using the expression

$$
\log BR \equiv \log 1/C = -k(\log P)^2 + k' \log P + k''
$$
 (17)

When electronic and steric effects in a set of congeners can be neglected, eq 17 often yields good correlations. For 16 different sets of hypnotics tested in a variety of ways on mice, rats, rabbits, and guinea pigs, an average value of log **Po** of *2* with a range of 1.5-2.7 was found. While log  $P_0$  says nothing about the intrinsic activity of a set of drugs, it does set a limit to the maximum activity one can expect to reach in a given set operating by a fixed mechanism by simply increasing the lipophilic character of the parent drug. It has been postulated<sup>24</sup> that there should be an ideal lipophilic character for a set of drugs so that the members having this value would be least restricted in their random walk through biological tissue to the sites of action. Once the drug molecule reaches the site of action, there is a last partitioning step onto a receptor site which may be highly important in determining BR and highly dependent on log P. To bring this out, eq 17 can be written<sup>24</sup> as:

$$
\log 1/C = -k_1 (\log P)^2 + k_2 \log P + k_3 \log P + k_4 \tag{18}
$$

In eq 18, k<sub>3</sub>log P comes from factoring the k<sup>'</sup>log P term of eq 17 to separate the dependence of log 1/C on the last partitioning step. If we assume that the hypnotics act by a rather nonspecific mechanism and the last partitioning step is little different from the many others, the empirically found log **Po** should be the same as the ideal value, log Pi, representing maximum freedom in the random walk process. There is some evidence that log Pi may be about *2.* The fact that benzeneboronic acids penetrating<sup>24</sup> into the mouse brain have log  $P_0$  of 2.3 (here the analytical tool is not biological response, but chemical analysis for boron) supports this view. **A** variety of other important, more or less *neutral,* CNS acting drugs<sup>24</sup> also have log P values near 2. The above concept removes part of the mystery of the so-called "blood-brain barrier." When active transport is not involved, the farther one is from log P of *2,* the lower the possibility a neutral drug has of penetrating the brain in a fixed time interval. When the  $k_3$ log P term in eq 18 becomes important, or when one has a situation approximating an equilibrium condition of the type envisaged by

**Chap. 33** Drug Design **Hansch** - **<sup>353</sup>** Ferguson, $^{26}$  then log P<sub>o</sub> will be different from log P<sub>i</sub>. Po for thiobarbiturates is found24 to be about **3.**  teria in *in vitro* tests, log Po is about *4,* and for Gram-positive bacteria it is found<sup>25</sup> to be about 6. Once one has determined log  $P_0$  for a given system by the testing of a few derivatives, considerable design work in the formulation of new derivatives for testing can be done by taking advantage of the additive-constitutive nature of  $\pi$  and log  $P^2$ ,  $2^8$ For example, log For Gram-negative bac-

Nonlinear Dependence of Activity on Electronic Factors - Just as departure from linearity between log 1/C and log P occurs when large changes are made in log P, so departure from linearity between log  $1/c$  and electronic changes results when large changes are made.  $\frac{11}{2}$ ,  $\frac{29}{2}$ ,  $\frac{30}{2}$  Two approaches have been applied to take account of this fact mathematically;  $Fujita^{11}$ , 29 has suggested the approach formulated in eq 19 for expressing activity in

$$
\log 1/C + \log(K_{\alpha} + [H^+]) / [H^+] = k\sigma + k'\pi + k''
$$
 (19)

terms of the unionized form of the drug. The left side of eq 19 can be replaced with log  $1/C + \log(K_a + [H^+] / K_a$  for the ionized form of the drug. Equations of the type of 19 have been shown to give good correlations with biologically-active phenols $^{29}$  and sulfa drugs. $^{11}\,$  However, as Fujita has pointed out,<sup>29</sup> considerable care must be exercised in interpreting such correlations. Since there is a direct relationship between  $\sigma$  and  $K_{a}$ , one in effect "builds in" correlation in this type of expression. Equation 19 does have the advantage that changes in the external pH are included, [H] The second approach to systems nonlinear in  $\sigma$  is to include a  $\sigma^2$  term. <sup>17,30</sup>

Higher Order Approximations - In the early attempts<sup>1</sup> to correlate structure with activity mathematically, simple linear combinations of physicochemical parameters were usually considered. It has become evident that the addition of interaction terms to such equations can in some instances yield sharper correlations.<sup>5,9,18,31</sup> Singer and Purcell<sup>31</sup> have discussed this problem and compared the models of Free and Wilson, Kopecky and Bocek and Hansch and his colleagues. They point out that in view of the many instances where BR is not linearly<sup>24</sup>,<sup>25</sup> related to log P and where BR is also not linearly related to electronic effects<sup>11,17,29,30</sup> that the model of Free and Wilson will not hold, but that the Kopecky-Bocek model<sup>32</sup> should apply.

Polarizability and Activity - Hersch,  $33$  in continuing his study of the re $lation ln(MBC) = ln C<sub>S</sub> - KR<sub>O</sub>I$ , has investigated the interaction of local anesthetics with lecithin monolayers. In the Hersch equation, **MBC** is the minimum concentration of drug blocking nerve excitability,  $C_S$  is the minimum blocking concentration of molecules at the surface,  $R_0$  is the mole refraction, I is the ionization potential, and K is a function of interaction distances which is assumed as a first approximation to be constant. Hersch has shown that a good linear relation exists between  $log(\triangle \pi / MBC)$ and RoI, where **An** represents change in surface pressure. Cammarata34 has analyzed the attempts to correlate structure with activity in chloramphenicol analogs. He has formulated eq 20 which gives a very high correlation with the single parameter  $P_E$ , molar electronic polarizability. The

$$
k_{\rm T} = 2.76P_{\rm T} - 6.55 \qquad \qquad n = 9 \qquad r = 0.991 \ (20)
$$

inhibitory constant,  $k_I$ , is from the work of Garrett. The results of Hersch and Cammarata indicate the importance of more exploratory work with  $P_{E}$ . This parameter might be particularly useful in separating charge transfer complexing ability of various functions from simple hydrophobic interactions.

Newer Semi-Empirical Approaches - Fuller, Marsh and Mills<sup>47</sup> have derived eq 21 which correlates inhibition of monoamine oxidase by N-(Phenoxyethyl)-cyclopropylamines. In eq 21,  $\gamma$  is an arbitrarily defined steric constant. Equation 21 was used to predict the activity of two new de-

$$
pI_{50} = 0.865\gamma + 0.209\pi + 1.547\sigma + 5.928 \qquad n = 16 \qquad r = 0.90 \quad (21)
$$

rivatives  $(4-N=N-C_6H_5$ ;  $4-NH_2$ ). Both derivatives were correctly predicted. One was found to be more active than any previously tested. The other had the expected low activity. Equation 21 was based on rat liver MAO. **A**  similar equation was formulated for inhibitors of the human enzyme. Turner and Battershell derived eq 22 which correlates the fungicidal activity of tetrachloroisophthalonitrile and its analogs in the foliage protectant

$$
\log C = -0.053(\log t_{1/2})^2 + 0.413 \log t_{1/2} + 0.513 \log VP + 1.021
$$

$$
n = 9
$$
  $r = 0.994 (22)$ 

test with early blight.<sup>48</sup> In eq 22, t<sub>1/2</sub> is the half reaction time of the fungicide with 4-nitrothiophenoxide and **VP** represents vapor pressure determined by GLC method. Since the fungicides were quite complex in structure, Turner and Battershell used the reaction parameter  $t_1/p$ . This parameter would of course contain both electronic and steric components. The thiophenoxide ion was used to represent a suspected enzymatic, **-SH,**  in the fungus. This appears to be an extremely useful general concept to employ when one is dealing with molecules too complex for M.O. calculated parameters or  $\sigma$  constants. McFarland  $e t$   $a l$ .<sup>49</sup> have derived eq 23 correlating structure and activity for pyrantel analogs. The addition

$$
log(1/MED) = -2.30\pi^2 + 2.83\pi - 0.320\sigma + 0.824\mu^2 + 0.988
$$

$$
n = 7
$$
  $r = 0.992$  (23)

of the dipole moment term **(p)** results in considerable improvement in correlation. Other studies on dipole moments $4l-44$  bring out the interest in the use of this constant for quantitative correlations. Purcell has extended his studies and shown the utility of eq 24 in the correlation of cholinesterase inhibition by  $3-(N, N-di)$ ethylcarbamoyl) piperidinoalkanes.<sup>43</sup>

$$
pI_{50} = \frac{n}{2} \left(\frac{1}{2.8}\right)^{n-1} A + Bn + C
$$
 (24)

Chap. **33** Drug Design Hansch - **<sup>355</sup>**

In eq 24,  $pI_{50}$  = -log  $I_{50}$  where I is the molarity causing 50% inhibition, n is the number of carbon atoms in an alkyl chain, A is an electronic factor, B a hydrophobic parameter, and C is the contribution to activity of the parent portion of the molecule. Amoore<sup>51</sup> continues to make progress in quantitatively correlating the shape of organic compounds with their odor. The similarity of molecules is compared by scanning their molecular silhouettes with a computerized pattern recognition machine.

Chromatographic Constants - Interest is developing<sup>1,48</sup> in the use of  $R_M$ values, obtained from paper or thin layer chromatography, to serve as a measure of hydrophobic character. Some evidence has been found<sup>1</sup> to show a linear relation between  $\pi$  and  $\Delta \mathbf{R_M}$ . Further evidence from the work of Bark and Graham<sup>35</sup> is contained in eq 25. In eq 25, R<sub>M</sub> values were ob-

 $R_M = 0.596\pi - 0.511$  n = 47  $r = 0.938$  (25)

tained from reversed phase thin layer chromatography using cellulose impregnated with ethyl oleate as the stationary phase. Aqueous ethanol (25% v/v) was used as the mobile phase. The  $\pi$  values are from the phenol system.15 Considering the movement of drugs through tissue as a kind of chromatographic process can lead to new insight in structure-activity study. For example, it is well known that stereoisomers move at different rates in chromatography. Even optical isomers can be separated when one employs an asymmetric adsorbent. This fact must be taken into account when one attempts to rationalize the difference in biological activity of different stereoisomers. At present, the almost universal tendency is to assume that differences in activity mean differences in fit to an asymmetric receptor site. **A** most important observation by Portoghese, Mikhail and Kupferberg<sup>50</sup> shows that the difference in brain concentration of epimeric analogs of meperidine is quantitatively related to their partition coefficients. This observation may well explain many examples that have been found where small to large differences in activity are recorded for different stereoisomers.

**M.O.** Parameters - Progress continues to be made in the correlation of pharmacological activity with electronic indices calculated from the molecular orbital approximation of quantum chemistry. **36-39** Purcell and Singer<sup>40</sup> have reviewed the parameters of use in the Hückel method. However, as has been previously noted,<sup>1</sup> it is rarely observed that quantitative equations such as those shown in this review result considering only M.O. parameters. Neely's work<sup>38</sup> does show that under certain limiting conditions such correlations can be found. The findings of Foernzler and Martin $^{37}$  are more typical. Although our understanding of freeenergy related extrathermodynamic relationships even in homogeneous systems is still far from complete<sup>52,53</sup> applications of important practical as well as theoretical value are being made in the area **of** medicinal chemistry.

## References



- 38. W. B. Neely, *Moz. PhannacoZ.,* **3,** 108 (1967).
- 39. K. Hirano, S. Yoshina, K. Okamura, and I. Suzuka, *Buzz. Chem. Soc. Japan,* 40, 2229 (1967).
- 40. W. P. Purcell and J. A. Singer, *J. Chern. Eng. Data,* 12, 235 (1967).
- 41. W. P. Purcell and J. A. Singer, *J. Phys. Chem.,* 71, 4316 (1967).
- 42. J. A. Singer, W. P. Purcell, and C. C. Thompson, *J. Med. Chem.,* 10, 528 (1967).
- $43.$ W. P. Purcell, Proc. Conf. Structure and Reactions of DFP Sensitive Enzymes, Stockholm, 1967, **p** 97.
- 44. E. J. Lien and W. D. Kumler, *J. Med. Chern.,* 11, 214 (1968).
- 45. R. Wildnauer and W. J. Canady, *Biochemistry,* 5, 2885 (1966).
- 46. R. Collander, *Acta Chem. Scand., 5,* 774 (1954).
- 47. R. W. Fuller, M. M. Marsh, and J. Mills, *J. Med. Chem.,* 11, 397 (1968).
- 48. N. J. Turner and R. D. Battershell, 155th National Meeting of the American Chemical Society, San Francisco, April, 1968.
- 49. J. W. McFarland, L. H. Conover, A. L. Howes, J. E. Lynch, 0. R. Chisholm, W. C. Austin, R. L. Cornwell, J. C. Danilewicz, W. Courtney, and D. H. Morgan, 155th National Meeting of the American Chemical Society, San Francisco, April, 1968.
- 50. P. S. Portoghese, A. A. Mikhail, and H. J. Kupferberg, *J. Med. Chern.,*  11, 219 (1968).
- 51. J. E. Amoore, G. Palmieri, and E. Wanke, *Nature,* 216, 1084 (1967).
- 52. C. D. Ritchie and W. F. Sager, *Progress in Physical Organic Chemistry,* 2, 323 (1964).
- 53. R. L. Schowen, *J. Phamn. Sci., 56,* 931 (1967).