hydration shell model are the direct result of independent theoretical calculations which have not been optimized on the basis of experimental data as have the π constants. This suggests that the theoretical calculations, and the associated potential energy functions, used to estimate the hydration shell parameters are realistic. However, these theoretical values of the hydration shell parameters could be used as initial parametric estimates from which an optimized set of values could be generated by using available experimental log (P_{wo})'s. Since interactions from second and, perhaps, even additional layers of solvent molecules probably are required to describe solvation thermodynamics, the differences between the optimized and existing hydrations shell parameters might be thought of as the contributions which the additional solvent layers make to the solute–solvent interactions.

The SCAP is limited to the calculation of dilute solution properties of a solute molecule. Solutions in which solute molecules associate, or the solute concentration is sufficiently high so as to alter the structure of bulk solvent, probably cannot be treated by SCAP. Still, it is our hope that this paper will be the first in a series reporting the successful application of SCAP to problems in molecular design. Our immediate interests will focus upon the prediction of optimim binding energy between some agonists and their receptors and its significance in generating useful QSAR.

Acknowledgment. A. J. Hopfinger expresses his thanks to Diamond Shamrock Corp. for support of this work. Both authors appreciate the helpful comment of Dr. N. Pollack of Diamond Shamrock Corp. during the preparation of this manuscript.

References and Notes

(1) Sloan Research Fellow.
(2) J. Ferguson, *Proc. R. Soc. London, Ser. B*, **127**, 387 (1939).
(3) A. Albert, "Selective Toxicity", Methuen, London, 1968, Chapter 14.
(4) For a review of the contributions by Hansch, see M. S. Tute, *Adv. Drug Res.*, **6**, 1–77 (1971).
(5) G. Redl, R. D. Cramer III, and C. E. Berkoff, *Chem. Soc. Rev.*, **3**, 273 (1974).
(6) L. B. Kier, "Molecular Orbital Theory in Drug Research", Academic Press, New York, N.Y., 1971.
(7) R. W. Taft, "Steric Effects in Organic Chemistry", Wiley, London, 1956.
(8) C. K. Hancock and C. P. Falls, *J. Am. Chem. Soc.*, **83**, 4214 (1961).
(9) E. Kutter and C. Hansch, *J. Med. Chem.*, **12**, 647 (1969).
(10) H. J. R. Weintraub and A. J. Hopfinger, *J. Theor. Biol.*, 41, 53 (1973).
(11) H. J. R. Weintraub and A. J. Hopfinger in "Molecular and Quantum Pharmacology", E. Bergmann and B. Pullman, Ed., Reidel, Dordrecht, Holland, 1974, p 131.
(12) A. J. Hopfinger, *Biopolymers*, **10**, 1299 (1971).
(13) F. R. Brown III, A. J. Hopfinger, and E. R. Blout, *J. Mol. Biol.*, **63**, 101 (1972).
(14) G. N. Ramachandran in "Peptides, Polypeptides, and Proteins", E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, Ed., Wiley-Interscience, New York, N.Y., 1974, p 1.
(15) H. A. Scheraga in ref 14, p 49.
(16) P. J. Flory in "Conformation of Biopolymers", G. N. Ramachandran, Ed., Academic Press, New York, N.Y., 1967, p 339.
(17) B. L. Farmer, A. J. Hopfinger, and J. B. Lando, *J. Appl. Phys.*, **43**, 4294 (1972).
(18) K. A. Maurtiz, E. Baer, and A. J. Hopfinger, *J. Polym. Sci., Polym. Phys. Ed.*, 11, 2185 (1973).
(19) A. J. Hopfinger, *Macromolecules*, **4**, 731 (1971).
(20) K. H. Forsythe and A. J. Hopfinger, *Macromolecules*, **6**, 423 (1973).
(21) A. J. Hopfinger in ref 14, p 71.
(22) A. J. Hopfinger, "Conformational Properties of Macromolecules", Academic Press, New York, N.Y., 1973.
(23) G. G. Nys and R. F. Rekker, *Chim. Ther.*, **8**, 521 (1973).
(24) G. G. Nys and R. F. Rekker, *Chim. Ther.*, **9**, 375 (1974).
(25) R. F. Rekker and G. G. Nys in ref 11, p 457.
(26) See Table I and references cited therein of C. Hansch, J. E. Quinlan, and G. L. Lawrence, *J. Org. Chem.*, **33**, 347 (1968).

Molecular Connectivity. 6. Examination of the Parabolic Relationship between Molecular Connectivity and Biological Activity

Wallace J. Murray,*[1] Lemont B. Kier,

Massachusetts College of Pharmacy, Boston, Massachusetts 02115

and Lowell H. Hall

Eastern Nazarene College, Quincy, Massachusetts 02169. *Received August 27, 1975*

The topologically derived, nonempirical molecular connectivity index, χ, for several classes of compounds is shown to be parabolically related to the biological activities of these compounds. Similar nonlinear relationships were previously shown between the octanol–water partition coefficients, expressed as log P, of the compounds and their biological activities. These and previous studies indicate that many physicochemical properties presently used in structure–activity studies may be intermediaries between the nonempirical molecular structure encoded in χ and measured biological activities.

The manner in which atoms in a molecule are arranged or connected is called molecular connectivity, a fundamental characteristic of any molecule. Many physical properties are dependent not only upon the number of atoms in a molecule (additive property) but also upon the arrangement of those atoms within the molecule (constitutive property). It is well established that simple lengthening of a hydrocarbon side chain by $-CH_2-$ units will correlate in a linear manner with most physical properties. However, branching in a hydrocarbon side chain has not previously been amenable to correlation. For example, pentyl alcohol, isopentyl alcohol, *sec*-pentyl alcohol, and *tert*-pentyl alcohol all have the same number of carbons and the same molecular weight but have different boiling points, molecular polarizabilities, solubilities, and partition coefficients.

For many years correlation of properties with branching has been attempted with but limited success.[2] Randic[3a] has recently formulated a branching index, later called χ or the molecular connectivity index,[3b] which seemingly encodes the additive and constitutive nature of a molecule. In the same way that the universal constant "π" is a

number showing a certain relationship in a circle, χ appears to be a number relating the shape and architecture of a molecule to the interatomic connections.

The size and shape of a molecule determine many physical properties, e.g., solubilities, boiling points, densities, heats of formation. In a series of five papers, Kier, Hall, and Murray[3b-7] have further developed and refined χ and have shown it to be significantly correlated to a number of physical properties which themselves are additive and constitutive. In addition, χ was shown to be correlated in a linear manner with a number of biological activities.[3b,5,6]

Many physicochemical properties are presently used in relating molecular structure to biological activities (structure–activity relationship or SAR). Recently, Norrington et al.[8] have suggested using the term "physicochemical–activity relationships" (PAR) to differentiate these methods from approaches which are strictly mathematical, e.g., the Free–Wilson method, or from approaches in which parameters are derived from theoretical calculations on molecular structures, e.g., quantum mechanical methods. Since physicochemical properties and not molecular structure per se are employed to explain the drug–biological interaction, adoption of the term, PAR, seems rational.

These properties, e.g., partition coefficients and molecular polarizabilities, are examined in multiple regression analyses to determine linear or curvalinear relationships between one or more of them and some measurable biological activity. Once a significant correlation has been found, certain conclusions are made concerning the fundamental events occurring in vivo. These physicochemical properties serve as models for biological processes which in turn are used as predictive devices for further design of improved or modified pharmaceutical agents.

The physicochemical properties have in common the fact that they must be experimentally determined. In some cases the additive nature of these properties is used to derive "predicted" properties. For example, Hansch and others[9,10] have made use of the additive nature of substituents on the log P of a parent molecule to derive substituent partition coefficients or π values. In fact, in most PAR studies employing log P, the values of the compounds studied are log P calculated from π values rather than the experimentally determined property.[9] Murray, Kier, and Hall[5] have shown that the additive nature of log P is a reflection of the more fundamental property of molecular connectivity as measured by χ.

The strength of χ lies in its strictly nonempirical derivation. It is calculated knowing only the molecular structure. In fact, it appears that in structure–activity relationships, one need not resort to physicochemical data such as log P or polarizabilities which are after all dependent on molecular structure. Rather, a correlation may be made directly between molecular structure as given by χ and biological activity. The experimentally derived data appear to be intermediates between fundamental structural characteristics and the biological activity being studied.

Previous papers have examined the linear relationship between χ and biological activities.[3,5,6] In this study we extend these relationships to the nonlinear case and look at the significance of these relationships in terms of the models used to explain the parabolic relationship between structure and biological activities.

Calculation of the Molecular Connectivity Index. The molecular connectivity index, χ, was calculated in the usual manner.[3] The molecular skeleton is drawn and each non-hydrogenic atom is assigned a number, δ, corresponding to the number of non-hydrogen atoms connected to it; thus, δ = 1, 2, 3, 4. A value for each bond in the molecule, C_k, is calculated from each pair of bonded atoms by $C_k = (\delta_i\delta_j)_k^{-1/2}$, bond k being that formed between atoms i and j. Finally, χ is calculated by summing C_k for all possible bonded atoms, $\chi = \Sigma C_k$. Cyclic compounds have one more bond than the corresponding straight-chained isomer and so the value of one ring of C_k is subtracted to arrive at χ.

Results and Discussion

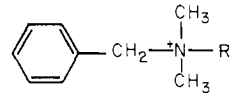
There are numerous examples showing parabolic relationships between a physicochemical property and some biological activity. In a recent review Hansch and Clayton[11] have compiled over 170 examples showing a nonlinear relationship between log P and biological activities. Lien[12] has also examined the parabolic relationship.

We have previously studied linear relationships between χ and biological activities, expressed as pC (–log $1/C$). This present study examines some of the nonlinear examples using the molecular connectivity index, χ, in place of the physicochemical property, log P, where P is the octanol–water partition coefficient. Tables I–V list these data. For Tables I–IV only sets having six or more data points were selected and only where the F test[13] indicated that the addition of the χ^2 term to the linear equation in χ was significant at the 0.99 level or better were they chosen. These data are based on homologues derived from changes in an aliphatic side chain. Table I illustrates the additivity principle in that for each increase in $-CH_2-$ unit there is an increase in χ of 0.5 units. The constitutive principle is illustrated by Tables II and III in that branching in the aliphatic chain also occurs. Prior to the introduction of the molecular connectivity index the structural effect of branching in a molecule could not be correlated to physicochemical or biological properties in a significant manner.[3a] The approach generally applied in estimating the structural effect of branching on properties has been to use a physicochemical property such as log P to correlate biological activities. This is what is meant when these properties are referred to as intermediaries between structure and activity.

Most examples showing parabolic relationships between structure and biological activity have modifications only in the hydrocarbon side chain. There are few studies wherein other groups such as –NH– or –O– are incorporated into the side chain as part of a systematic analogue study.[11] In their review Hansch and Clayton have compiled a list of 173 studies showing parabolic behavior.[11] In only seven of these studies are there modifications other than in the hydrocarbon side chain. And in these seven examples a hydrocarbon side chain is usually varied along with the other moiety change.

Table IV presents a study wherein two systemic variations are being made. One variation incorporates group changes in the ring system (H–, Cl–, CH_3O-, and $-OCH_2O-$). The other modification is with the hydrocarbon side chain of a quaternary ammonium group. The biological activity being measured is the minimum killing concentration (MKC) on *Staphylococcus typhosa* and *Staphylococcus aureus*.[14] The authors have made linear correlations with critical micelle formation and the number of carbons on the quaternary nitrogen (the additive principle). They attribute the parabolic behavior of antibacterial action to the effect of two competing processes. One is the attraction of the hydrophobic portion of the molecule to the bacterial surface. The other is the

Table I. Molecular Connectivity Index vs. Minimum Inhibitory Concentration (MIC), Minimum Killing Concentration (MKC), or Minimum Concentration for 50% Hemolysis ($C_{H_{50}}$) of a Congeneric Series of Quaternary Ammonium Salts[a]

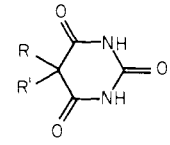


pC

		I-A		I-B		I-C		I-D		I-E		I-F		I-G	
R	χ	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd
C_8H_{17}	8.306			3.69	3.660	3.00	2.682	1.68	1.482	2.69	2.512	1.52	1.666	1.76	1.736
C_9H_{19}	8.806	2.54	2.414	4.20	4.263	3.02	3.280	2.08	2.141	3.02	3.165				
$C_{10}H_{21}$	9.306	2.74	2.862	4.74	4.766	3.57	3.793	2.41	2.683	3.57	3.724	3.45	3.209	2.95	2.964
$C_{11}H_{23}$	9.806	3.06	3.223	5.06	5.168	4.06	4.221	2.92	3.108	4.06	4.189				
$C_{12}H_{25}$	10.306	3.61	3.495	5.61	5.469	4.61	4.563	3.45	3.417	4.61	4.561	4.34	4.266	3.82	3.861
$C_{13}H_{27}$	10.806	3.80	3.679	5.80	5.670	5.10	4.821	3.85	3.609	5.10	4.839				
$C_{14}H_{29}$	11.306	3.65	3.774	5.82	5.771	5.12	4.993	3.96	3.684	5.12	5.023	4.63	4.837	4.40	4.430
$C_{15}H_{31}$	11.806	3.84	3.782	5.84	5.770	5.14	5.080	3.74	3.643	5.14	5.113				
$C_{16}H_{33}$	12.306	3.68	3.702	5.56	5.668	5.16	5.082	3.30	3.486	5.16	5.110	4.88	4.922	4.79	4.669
$C_{17}H_{35}$	12.806	3.57	3.534	5.18	5.467	4.70	4.999	3.06	3.212	4.70	5.013				
$C_{18}H_{37}$	13.306			5.19	5.164	4.71	4.830	2.63	2.821	4.71	4.822	4.60	4.520	4.52	4.579
$C_{19}H_{39}$	13.806	2.91	2.934	4.91	4.760	4.73	4.576	2.52	2.314	4.73	4.538				

[a] Data from ref 10.

Table II. Molecular Connectivity Index vs. Barbiturate Activity[a]



pC

			II-A		II-B		II-C		II-D	
R	R′	χ^c	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd
Methyl	$CH_3CHC(CH_3)_3$	7.097			2.64	2.676				
Methyl	$CH_2C(CH_3)_3$	6.649			2.12	2.146				
Ethyl	Ethyl	6.064	3.09	3.046			2.91	2.880	1.32	1.344
Ethyl	Isopropyl	6.446	3.30	3.380			3.34	3.282	1.89	1.927
Ethyl	Butyl	7.064	3.72	3.648			3.53	3.580	2.80	2.626
Ethyl	Isobutyl	6.920	3.63	3.615						
Ethyl	*sec*-Butyl	6.984	3.63	3.632						
Ethyl	*sec*-Pentyl	7.484							3.07	2.928
Ethyl	Isopentyl	7.457	3.75	3.644			3.59	3.543	3.12	2.913
Ethyl	Hexyl	8.063							3.40	3.117
Ethyl	Allyl	6.917					3.28	3.373		
Ethyl	$CH_2C(CH_3)_3$	7.210			2.91	2.785				
Ethyl	$CH_3CH(CH_3)_3$	7.658			3.15	3.122				
Ethyl	Phenyl	7.961	3.46	3.541					2.36	3.056
Propyl	Propyl	7.063	3.55	3.647						
Propyl	Isopropyl	6.946	3.63	3.622						
Propyl	Isopentyl	7.920	3.48	3.465						
Propyl	$CH_3CHC(CH_3)_3$	8.158			3.29	3.315				
Propyl	$CH_2C(CH_3)_3$	7.710			3.04	3.151				
Propyl	Allyl	7.417					3.47	3.580		
Isopropyl	Butyl	7.446					3.49	3.547		
Isopropyl	Allyl	7.234					3.60	3.557		
Butyl	Butyl	8.064					3.08	3.139		
Butyl	Allyl	7.917					3.47	3.503		
Butyl	$CH_3CHC(CH_3)_3$	8.658			3.36	3.316				
Butyl	$CH_2C(CH_3)_3$	8.210			3.33	3.324				
Isobutyl	Allyl	7.773					3.63	3.554	2.80	2.891
Isobutyl	$CH_2C(CH_3)_3$	8.066			3.27	3.294				
sec-Butyl	Allyl	7.734					3.78	3.534		
Pentyl	$CH_2C(CH_3)_3$	8.710			3.32	3.304				
Isopentyl	Allyl	8.273					3.45	3.504		
Isopentyl	$CH_2C(CH_3)_3$	8.566			3.26	3.330				
Allyl	*sec*-Pentyl	8.234							3.19	3.107
Allyl	Cyclopentyl	7.608							2.90	2.940
Allyl	Allyl	7.063					3.54	3.580		
Allyl	$CH_3CH(CH_3)_3$	8.446			3.39	3.315				

[a] Data from ref 10.

tendency for the molecules to cling together as the hydrocarbon chain grows larger.

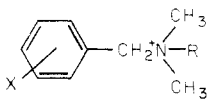
Table V shows another example of dual variation. In addition to changes occurring in a hydrocarbon side chain,

Table III. Molecular Connectivity Index vs. in Vitro Intestinal Absorption Data for Congeneric Series of Carbamates[a]

		Log (% A/cm²/h)					
		Serosal transfer, III-A		Tissue bound, III-B		Mucosal loss, III-C	
R	χ	Obsd	Calcd	Obsd	Calcd	Obsd	Calcd
Methyl	2.270	0.45	0.420	−0.38	−0.405	0.52	0.543
Ethyl	2.770	0.53	0.609	−0.22	−0.176	0.60	0.691
Propyl	3.270	0.71	0.694	−0.04	−0.030	0.78	0.749
tert-Butyl	3.417	0.65	0.699	0.00	−0.003	0.73	0.748
Isobutyl	3.626	0.75	0.690	0.04	0.024	0.83	0.735
Butyl	3.770	0.82	0.674	0.10	0.033	0.90	0.717
tert-Pentyl	3.917	0.68	0.648	0.04	0.036	0.77	0.690
Pentyl	4.270	0.57	0.550	0.02	0.014	0.68	0.594
tert-Hexyl	4.417	0.09	0.494	−0.21	−0.007	0.26	0.541
Hexyl	4.770	0.35	0.322	−0.05	−0.088	0.50	0.382
Heptyl	5.270	−0.17	−0.010	−0.28	−0.272	0.08	0.079
Octyl	5.770	−0.43	−0.447	−0.57	−0.539	−0.19	−0.314
Benzyl	4.900	0.59	0.246	0.01	−0.127	0.69	0.312

[a] Data from ref 12.

Table IV. Molecular Connectivity Index vs. Minimum Killing Concentration (MKC) of a Series of Substituted Aromatic Quaternary Ammonium Salts[a]



			pC	
X	R	χ	Obsd	Calcd
H	$C_{10}H_{21}$	9.138	2.79	3.02
H	$C_{12}H_{25}$	10.138	3.74	3.77
H	$C_{14}H_{29}$	11.138	4.17	4.13
H	$C_{16}H_{33}$	12.138	3.92	4.09
H	$C_{18}H_{37}$	13.138	3.34	3.65
2-Cl	C_8H_{17}	8.549	2.02	2.39
2-Cl	$C_{10}H_{21}$	9.549	3.41	3.37
2-Cl	$C_{12}H_{25}$	10.549	4.14	3.96
2-Cl	$C_{14}H_{29}$	11.549	4.14	4.16
2-Cl	$C_{16}H_{33}$	12.549	3.74	3.95
2-Cl	$C_{18}H_{37}$	13.549	3.43	3.35
4-Cl	C_8H_{17}	8.532	2.60	2.37
4-Cl	$C_{10}H_{21}$	9.532	3.60	3.36
4-Cl	$C_{12}H_{25}$	10.532	4.04	3.96
4-Cl	$C_{14}H_{29}$	11.532	4.34	4.16
4-Cl	$C_{16}H_{33}$	12.532	3.85	3.96
4-Cl	$C_{18}H_{37}$	13.532	3.17	3.37
2-Cl, 4-Cl	C_8H_{17}	8.943	2.63	2.83
2-Cl, 4-Cl	$C_{10}H_{21}$	9.943	3.65	3.66
2-Cl, 4-Cl	$C_{12}H_{25}$	10.943	4.25	4.09
2-Cl, 4-Cl	$C_{14}H_{29}$	11.943	4.28	4.13
2-Cl, 4-Cl	$C_{16}H_{33}$	12.943	3.41	3.76
2-Cl, 4-Cl	$C_{18}H_{37}$	13.943	3.30	3.01
3-Cl, 4-Cl	C_8H_{17}	8.943	2.92	2.83
3-Cl, 4-Cl	$C_{10}H_{21}$	9.943	3.79	3.66
3-Cl, 4-Cl	$C_{12}H_{25}$	10.943	4.36	4.09
3-Cl, 4-Cl	$C_{14}H_{29}$	11.943	4.04	4.13
3-Cl, 4-Cl	$C_{16}H_{33}$	12.943	3.39	3.76
3-Cl, 4-Cl	$C_{18}H_{37}$	13.943	3.11	3.01
3-OCH_3, 4-OCH_3	$C_{12}H_{25}$	12.019	3.71	4.11
3-OCH_3, 4-OCH_3	$C_{14}H_{29}$	13.019	4.11	3.72
3-OCH_3, 4-OCH_3	$C_{16}H_{33}$	14.019	3.41	2.93
3,4-(OCH_2O)	C_8H_{17}	8.288	2.04	2.07
3,4-(OCH_2O)	$C_{10}H_{21}$	9.288	3.08	3.16
3,4-(OCH_2O)	$C_{12}H_{25}$	10.288	4.00	3.85
3,4-(OCH_2O)	$C_{14}H_{29}$	11.288	4.23	4.15
3,4-(OCH_2O)	$C_{16}H_{33}$	12.288	4.20	4.05
3,4-(OCH_2O)	$C_{18}H_{37}$	13.288	3.23	3.55

[a] Data from ref 10.

in this case an ether group, the ring system is also varied (anilines, aminopyridines, aminopyrimidines, and naphthylamines). The biological action examined was the minimum inhibitory concentration (MIC) on *Mycobacterium tuberculosis*. But again the data indicate that

Table V. Molecular Connectivity Index vs. Minimum Inhibitory Concentration of a Series of Aromatic and Heterocyclic Amines[a]

		pC	
Compound[b]	χ	Obsd	Calcd
4-OCH_3-An	3.826	3.39	2.98
4-OC_2H_5-An	4.326	4.44	4.06
4-OC_4H_9-An	5.326	5.42	5.26
4-OC_5H_{11}-An	5.826	5.45	5.38
4-OC_6H_{13}-An	6.326	5.80	5.17
4-OC_8H_{15}-An	7.326	3.44	3.79
2-OCH_3-3-NH_2-Py	3.842	2.59	3.02
2-OC_2H_5-3-NH_2-Py	4.342	3.54	4.09
2-OC_3H_7-3-NH_2-Py	4.842	4.79	4.84
2-*i*-OC_4H_9-3-NH_2-Py	5.198	4.52	5.18
2-OC_5H_{11}-3-NH_2-Py	5.342	5.76	5.27
2-OC_7H_{15}-3-NH_2-Py	6.342	4.62	5.16
2-OC_8H_{17}-3-NH_2-Py	6.842	3.75	4.62
2-$OC_{10}H_{21}$-3-NH_2-Py	7.842	3.19	2.57
4-OC_4H_9-3-NH_2-Py	5.342	5.73	5.27
4-OC_6H_{11}-3-NH_2-Py	6.342	5.80	5.16
2-OC_4H_9-5-NH_2-Pyr	5.326	5.13	5.26
2-OC_6H_{13}-5-NH_2-Pyr	6.326	5.49	5.17
4-OC_4H_9-1-NA	6.826	4.03	4.64

[a] Data from ref 10. [b] An = aniline; Py = pyridine; Pyr = pyrimidine; NA = naphthylamine.

parabolicity can be attributed to the lengthening of the size of the hydrocarbon side chain.

Regression analyses show the same level of significance for χ^2 as for $(\log P)^2$. Table VI summarizes the nonlinear equations for each example and in Table VII is compiled a comparison of the statistics obtained from the regression analyses for $\log P$ and for χ. In each case r, the correlation coefficient, and s, the standard deviation, have very similar values for expressions utilizing χ as those utilizing $\log P$.

What have we gained from such a study? As pointed out, in most studies employing $\log P$, the $\log P$ values are generally calculated utilizing the additive relationships thought to exist.[9] The π values for substituents have been compiled in tables for such parent compounds as phenol, phenoxyacetic acid, aniline, etc. To synthesize a $\log P$, one chooses an appropriate parent molecule and then adds up π values for each of the substituents attached to it. A problem may arise if one has a complicated molecule as to which parent system to select. Another problem arises because more than one partitioning system has been used to compile $\log P$ values and $\log P$ varies according to partitioning system as well as experimental conditions.[12]

No such problems exist in utilizing the molecular connectivity index, χ. The additive nature of $\log P$ is well

Table VI. Nonlinear Equations for Data in Tables I-V

Eq no.	$pC = a\chi^2 + b\chi + c$ a	b	c	Data ref	Biological act.[a]
1	−0.176 ± 0.0173	4.086 ± 0.390	−19.91 ± 2.16	I-A	*P. aeruginosa*, MIC
2	−0.201 ± 0.0154	4.654 ± 0.341	−21.10 ± 1.85	I-B	*S. aureus*, MIC
3	−0.170 ± 0.0252	4.111 ± 0.559	−19.71 ± 3.04	I-C	*Cl. welchii*, MIC
4	−0.233 ± 0.0240	5.306 ± 0.533	−26.51 ± 2.90	I-D	*P. aeruginosa*, MKC
5	−0.187 ± 0.0208	4.512 ± 0.460	−22.03 ± 2.50	I-E	*Cl. welchii*, MKC
6	−0.243 ± 0.0348	5.823 ± 0.755	−29.93 ± 4.00	I-F	Red cell sheep, $C_{H_{50}}$
7	−0.165 ± 0.0139	4.126 ± 0.300	−21.18 ± 1.59	I-G	Red cell sheep, $C_{H_{50}}$
8	−0.439 ± 0.0715	6.361 ± 1.007	−19.39 ± 3.53	II-A	Rabbit, MHD
9	−0.385 ± 0.0534	6.473 ± 0.828	−23.88 ± 3.19	II-B	Mouse, AD_{50}
10	−0.571 ± 0.0843	8.193 ± 1.188	−25.82 ± 4.16	II-C	Rat, MHD
11	−0.395 ± 0.273	6.469 ± 3.856	−23.36 ± 13.54	II-D	Brain rat, I_{50}
12	−0.209 ± 0.0495	1.429 ± 0.403	−1.749 ± 0.791	III-A	Intestinal absorption rates
13	−0.165 ± 0.0223	1.291 ± 0.181	−2.484 ± 0.356	III-B	Intestinal absorption rates
14	−0.180 ± 0.0435	1.206 ± 0.354	−1.265 ± 0.696	III-C	Intestinal absorption rates
15	−0.198 ± 0.014	4.577 ± 0.322	−22.24 ± 1.78	IV	*S. aureus*, MKC
16	−0.645 ± 0.091	7.421 ± 0.043	−15.98 ± 2.91	V	*M. tuberculosis*, MIC

[a] Abbreviations for biological activities and systems include MIC (minimum inhibitory concentration); MKC (minimum killing concentration); $C_{H_{50}}$ (minimum concentration for 50% hemolysis); MHD (minimum hypnotic dose); AD_{50} (minimum anesthetic dose for 50% of the population); I_{50} (concentration for 50% inhibition).

Table VII. Comparison of Statistics for Equations Using $(\log P)^2$ vs. Equations Using χ^2

Eq no.	n	Statistics for $(\log P)^2$ r	s	Statistics for χ^2 r	s
1	10	0.974	0.123	0.974	0.123
2	12	0.979	0.158	0.982	0.141
3	12	0.966	0.230	0.966	0.230
4	12	0.962	0.220	0.962	0.220
5	12	0.980	0.190	0.980	0.190
6	6	0.991	0.213	0.991	0.213
7	6	0.998	0.084	0.998	0.086
8	12	0.973	0.096	0.942	0.077
9	14	0.884	0.115	0.986	0.070
10	10	0.991	0.076	0.914	0.100
11	13	0.841	0.100	0.906	0.311
12	13	0.860	0.080	0.894	0.185
13	13	0.929	0.130	0.924	0.084
14	13	0.888	0.067	0.904	0.163
15	38	0.876	0.304	0.931	0.230
16	19	0.895	0.487	0.872	0.535

described by χ as shown by its significant correlation to the more fundamental property of molecular connectivity.[5] The calculation of χ is not dependent upon any experimental system or conditions. It is not dependent upon a compilation of experimental values in a table nor upon the selection of a correct parent molecule. It is derived only from the molecular framework by simple arithmetic processes and is, in that sense, nonempirical.

Parabolic relationships have been explained in many ways.[11,12,15] If a significant correlation is found between $(\log P)^2$ and biological activity, the usual explanation given is based on the partitioning model. In this model the drug molecule is pictured as partitioning on and off lipophilic macromolecules or distributing in and out of body "compartments" in a random walk process. However, examination of examples showing parabolic relationships reveals that the majority involve monocellular or monocomponent processes, e.g., antibacterial activities, an isolated animal or plant cell system, mitochondrial inhibition, or transport across membranes. Another large category is enzymatic activities, e.g., monoamine oxidase. Only a few examples are whole body studies. It seems difficult to rationalize a partitioning model occurring in the majority of these systems, especially when the

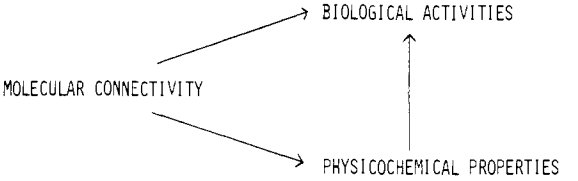


Figure 1. Relationship of molecular connectivity to physicochemical properties and biological activities.

mathematical model indicates that transport across more than ten barriers is necessary to show a nonlinear relationship.[16]

Alternative mechanistic explanations for parabolic relationships have been given.[11] They include (1) the principle of bulk tolerance, (2) limited solubility of the higher members of a congeneric series, (3) conformational distortion of the active site, and (4) metabolic transformations. All of these explanations predict that molecular size is the governing influence. The molecular connectivity index, χ, which mirrors molecular connectedness is most readily identified with the size and shape of a molecule. The good correlation between χ and biological activity makes the principle of bulk tolerance an attractive mechanistic model. This model predicts that in a congeneric series a point is reached wherein it becomes more difficult for each successively larger derivative to fit onto or into the active site.

χ has been correlated to solubility[4] so that the limited solubility explanation given by Ferguson[17,18] would also be a plausible model. One might view this explanation as comprising two or more competing processes. For example, the data presented in Table IV are explained in terms of the competition between the affinity of the drug molecule for the bacterial surface and the affinity for itself which would limit the number of drug molecules available to the bacterial wall.[14]

There may be many reasons for parabolic behavior. To choose any one mechanism or combination of mechanisms for a given situation would be difficult and may be an unwarranted simplification. The use of χ does not necessarily call a specific mechanism into play.

Figure 1 shows how the relationship between molecular connectivity and physicochemical properties and biological activities might be pictured. The index, χ, reflects the

additive and constitutive nature of molecules, i.e., the molecular size and shape. It has been significantly correlated to many physicochemical properties used in physicochemical–activity studies. If a biological response of a series of closely related compounds is significantly correlated to χ it is mirroring the same additive and constituitive molecular property shown in those physical properties studied thus far. Logically, it would appear that these empirical properties are intermediary measures of the very fundamental molecular structure encoded in χ and the measured biological response.

References and Notes

(1) Address correspondence to this author at the College of Pharmacy, University of Nebraska Medical Center, Lincoln, Neb. 68588.
(2) D. H. Rouvray, *Am. Sci.*, **61**, 729 (1973).
(3) (a) M. Randić, *J. Am. Chem. Soc.*, **97**, 6609 (1975); (b) L. B. Kier, L. H. Hall, and W. J. Murray, *J. Pharm. Sci.*, **64**, 1971 (1975).
(4) L. H. Hall, L. B. Kier, and W. J. Murray, *J. Pharm. Sci.*, **64**, 1974 (1975).
(5) W. J. Murray, L. B. Kier, and L. H. Hall, *J. Pharm. Sci.*, **64**, 1978 (1975).
(6) L. B. Kier, W. J. Murray, and L. H. Hall, *J. Med. Chem.*, **18**, 1272 (1975).
(7) L. H. Hall, L. B. Kier, and W. J. Murray, *J. Phys. Chem.*, manuscript submitted for publication.
(8) F. E. Norrington, R. M. Hyde, S. G. Williams, and R. Wootton, *J. Med. Chem.*, **18**, 604 (1975).
(9) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71**, 525 (1971).
(10) C. Hansch and W. J. Dunn, III, *J. Pharm. Sci.*, **61**, 1 (1972).
(11) C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1973).
(12) E. J. Lien in "Medicinal Chemistry", Vol. IV, J. Maas, Ed., Elsevier, Amsterdam, Netherlands, 1974.
(13) N. R. Draper and H. Smith, "Applied Regression Analysis", Wiley, New York, N.Y., 1966.
(14) S. Ross, C. E. Kwartler, and J. H. Bailey, *J. Colloid Sci.*, **8**, 385 (1953).
(15) T. Higuchi and S. S. Davis, *J. Pharm. Sci.*, **59**, 1376 (1970).
(16) J. T. Penniston, L. Beckett, D. L. Bently, and C. Hansch, *Mol. Pharmacol.*, **5**, 333 (1969).
(17) J. Ferguson, *Proc. Roy. Soc. (London), Ser. B*, **127**, 387 (1939).
(18) J. Ferguson, *Chem. Ind. (London)*, 818 (1964).

Quantitative Structure–Activity Relationships. 1. The Modified Free–Wilson Approach

Hugo Kubinyi* and Ott-Hermann Kehrhahn

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

Received June 27, 1975

The relationships between the linear free energy related Hansch model and the mathematical models of Free–Wilson and Bocek–Kopecký are reviewed and discussed. Some examples are given to illustrate the theoretically derived relationships and to demonstrate scope and limitations of each mathematical model. The modified Free–Wilson approach is shown to be completely equivalent to a nonparabolic Hansch approach; it can be used to study additivity or nonadditivity of group contributions and to control and improve the fitting of Hansch equations. The Bocek–Kopecký approach is related to the parabolic form of the Hansch approach; its practical use is limited by the great number of variables involved.

In 1964 three different approaches for studying quantitative structure–activity relationships were developed: Hansch's linear multiple regression model[1-3] (e.g., eq 1),

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

Free–Wilson's additive model[4] (eq 2), and Bocek–

$$\text{biological act.} = \text{overall av} + \Sigma(\text{group contributions}) \quad (2)$$

Kopecký's interaction model[5,6] (eq 3).

$$\log 1/C = b_X + b_Y + e_Xe_Y + k \quad (3)$$

Free and Wilson's model is based on the assumption that each substituent makes an additive and constant contribution to the biological activity regardless of substituent variation in the rest of the molecule. The values of the individual group contributions are calculated by regression analysis (an excellent introduction into Hansch and Free–Wilson analysis for people not familiar with mathematics is given in ref 7).

Bocek and Kopecký's interaction model may be interpreted as a Free–Wilson-like additive model with an additional term e_Xe_Y accounting for possible interactions between substituents X and Y. Although Bocek–Kopecký's model is cited in almost every review article on quantitative structure–activity relationships, it has found no practical use due to the great number of parameters involved.

Singer and Purcell[8] studied the relationships among the linear free energy based Hansch approach and the two mathematical models. They could demonstrate that all models are theoretically interrelated but Free–Wilson's model is appropriate only in the case of additivity of group contributions while Bocek–Kopecký's interaction model also holds in the case of parabolic dependence of biological activity on a particular physical property, e.g., Hansch's substituent constant π. In view of these relationships between the mathematical models and the Hansch approach they proposed the use of log $1/C$ values instead of linear values[4] as biological response parameters in Free–Wilson analysis. In the following years there has been some discussion whether log $1/C$ or C should be used,[9,10] but today in most instances log $1/C$ is being used. It should be noted that Bruice et al.[11] were the first ones who used log $1/C$ values and an additive model to calculate the activity of thyroxine analogues (eq 4).

$$\log \%\ \text{thyroxine-like activity} = k\Sigma f + c \quad (4)$$

Cammarata and Yau[12] and Fujita and Ban[13] used a modified Free–Wilson approach (eq 5)

$$\log 1/C = \sum_i a_i + \mu \quad (5)$$