LBL-6372
c. 1

Presented at the Conference on Water Chlorination: Environmental Impact and Health Effects, Gatlinburg, TN, October 31 - November 4, 1977.

Energy and Environment Division

A Priori Predictive Methods Of Assessing Health Effects Of Chemicals In The Environment

Mathilde J. Kland

October 1977

Lawrence Berkeley Laboratory University of California/Berkeley

Prepared for the U.S. Department of Energy under Contract No. W-7405-ENG-48

LBL-6372
c. 1

A PRIORI PREDICTIVE METHODS OF ASSESSING HEALTH EFFECTS OF CHEMICALS IN THE ENVIRONMENT*

Mathilde J. Kland

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

October 1977

*Work supported by U.S. Department of Energy

This report was done with support from the United States Energy Research and Development Administration. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the United States Energy Research and Development Administration.

INTRODUCTION

Passage of the Toxic Substances Control Act (TSCA)[1] last year emphasized the urgent need for the formulation of viable criteria and interim standards limiting the exposure of increasingly large segments of the U.S. population to environmental chemical toxicants. Unfortunately, current methods of developing these standards are both time-consuming and costly.[2] The resulting need for a priori predictive techniques to assess the inherent potential of chemicals, such as the halocarbons found in chlorinated waters, for inducing adverse biological effects, has led to the use of a number of analytical methods designed primarily for screening large numbers of chemical compounds before they impose unacceptable environmental hazards, frequently of crisis proportions.[3-5]

Four of the techniques best adapted to dealing with the multifactorial environmental problems of chemical health effects will be briefly described: (1) quantitative structure/activity relationships (QSAR); (2) factor analysis (FA); (3) pattern recognition/artificial intelligence (PR/AI); and (4) molecular connectivity (MC).

Historical

About 1870 Crum Brown and Fraser[6] enunciated the basic relationship between biological response (R) and drug structure (C),

$$R = f(C) \tag{1}$$

thus laying the groundwork for the subsequent structure-activity relationship (SAR) investigations of Hansch and others.* The quantitative foundations of SAR have their origins in Hammett's studies of reaction rates and equilibrium constants for series of derivatives of benzoic acids[8a,b] from which he deduced the well-known relationship:

$$\log K_\sigma = \log K_u + \alpha\rho \tag{2}$$

where σ is a constant characteristic of the substituent and its location, ρ a (normalized) reaction constant, and K_σ and K_u the corresponding rate or equilibrium constants for the substituted and unsubstituted species, respectively.

---

* See Ref. 7 for an excellent historical and state-of-the-art review of QSAR.

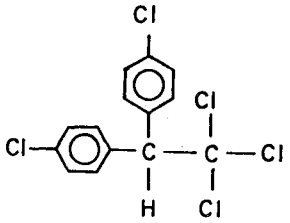
Hansch et al.[9a,b] used an analogous substituent constant π, where

$$\pi = \log P_X - \log P_H \quad , \tag{3}$$

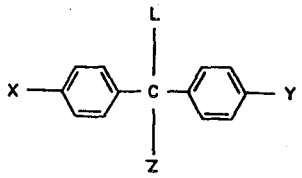
expressed in terms of the partition coefficients in octanol/water of parent compounds (P_H) and their substituted derivatives (P_X). With the π function, linear regression analyses were developed for substituent effects on the biological activities of benzoic acids on mosquito larvae; phenols on gram-positive and gram-negative bacteria; phenylethyl phosphate insecticides on houseflies; thyroxine derivatives on rodents; dimethylaminoethyl benzoates on guinea pigs; and carcinogenic compounds on mice. Clearly, this was a predictive tool of considerable potential.

SAR and the DDT Problem

In the course of a systematic search for general contact poisons for insect pests, Paul Müller "rediscovered" Dichlorodiphenyl Trichloroethane (DDT,I).*



(I) DDT



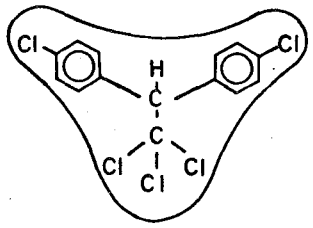
(II) DDT analog

The economic importance of DDT and the need for environmentally less-persistent substitutes, have resulted in many studies involving SAR and molecular modeling,[10a,b] degradation,[11] and immunological investigations.[12a,b] Fahmy et al.[10b] used the generalized model of a DDT analog (II).

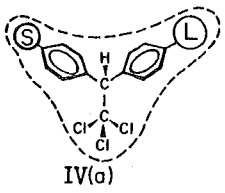
It was assumed that DDT and its analogs fit into the receptor site (R) of a macromolecule (e.g., a protein or lipoprotein in a nerve membrane), as shown in structure III. For optimum interaction the integral size of the whole molecule

---

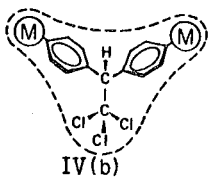
*Synthesized at J.R. Geigy, A.G., Switzerland (September 25, 1939).



(III)



IV(a)



IV(b)

$$\text{Log } LD_{50} = \alpha + \beta E_s^X + \gamma [E_s^X]^2$$

is important. This is illustrated in Structures IVa and IVb, where M is a substitute similar in size to chlorine; group S is smaller and L is larger than M. It is clear that the R-site can accommodate unsymmetrically substituted analogs, provided that the integral size of the molecule can fit within the flexible boundary of the site. Further, the model implies that as the sums of groups X, Y, L, and Z increase, interaction with R will increase to a maximum and thereafter decrease when the limit for optimum interaction (best fit) is exceeded. Because it was likely that steric effects would be a dominant determinant in fit at the R-site, the Taft steric substituent parameter (E_s)[13] was selected for estimating substituent size. The model implied a parabolic relationship of E_s and toxicity. Thus, for a single substituent X, the equation was of the form:

$$\log LD_{50} = \alpha + \beta E_s^X + \gamma [E_s^X]^2 \quad , \qquad (4)$$

where α, β, and γ are constants. This model appears to offer a reasonable explanation for the toxic behavior of DDT analogs with one smaller and one larger substituent than chlorine in the X and Y positions as illustrated in Fig. 1 with a plot of synergized log LD_{50} for houseflies against E_s for different substituents in position X while Y is held constant (Y = Me). Except for the identity case (X = Y = Me), the plot is remarkably similar to a potential energy diagram for diatomic molecules. It indicates an optimum E_s region where maximum interaction between chemical compound and the R-site takes place.

The concept of molecular size and fit has been used extensively by Metcalf and others in their design of DDT-like insecticides with more polar, less environmentally persistent characteristics. Similar studies are needed relating E_s and other parameters to the toxicities of those halo-organics produced during water chlorination.

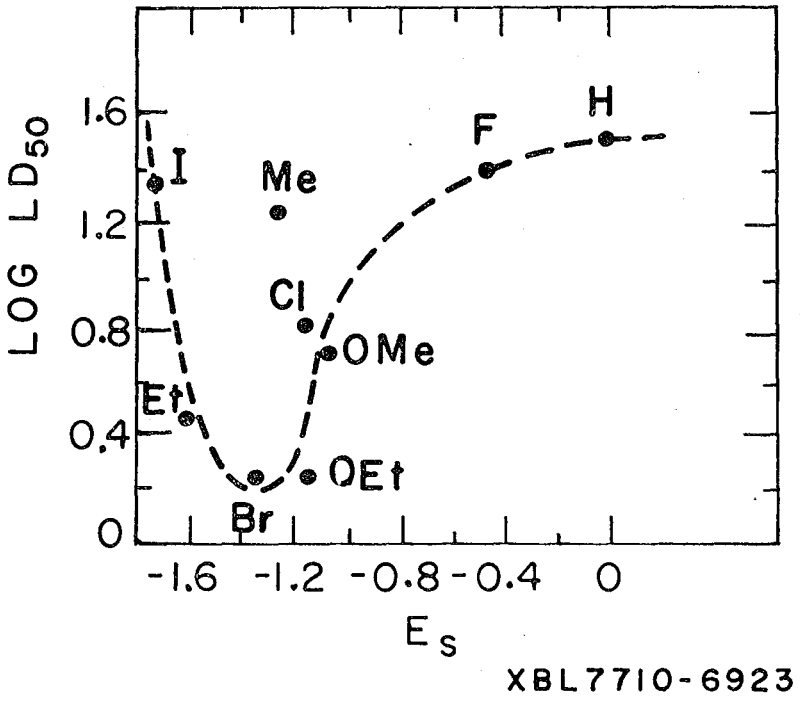


Figure 1. Relationship between toxicity and E_s for 1,1,1-trichloro-p-methyl-p'-x-diphenylethanes.

FACTOR ANALYSIS (FA)

Factor analysis (FA) is one of a number of statistical methods which can cope with large quantities of data of the kind encountered in dealing with interactions in chemical and biological systems. Originally applied largely in the social sciences,[14] FA has in recent years found increasing application to chemical[15,16a-c] and biological[17a,b] problems.

Some Basic Assumptions of FA

A detailed formulation of the technique of FA is clearly beyond the scope of this paper. However, a summary of the basic principles applied to simple chemical systems follows.

For a two-dimensional data matrix, two mathematical requirements must be satisfied by the property measured. First, each data point (D) is expressed as a linear sum of terms:

$$D = d_1 + d_2 + ... + d_n . \quad (6)$$

Second, each data point D is also a sum of row and column product terms,

$$D = r_1c_2 + r_1c_2 + ... + r_nc_n , \quad (7)$$

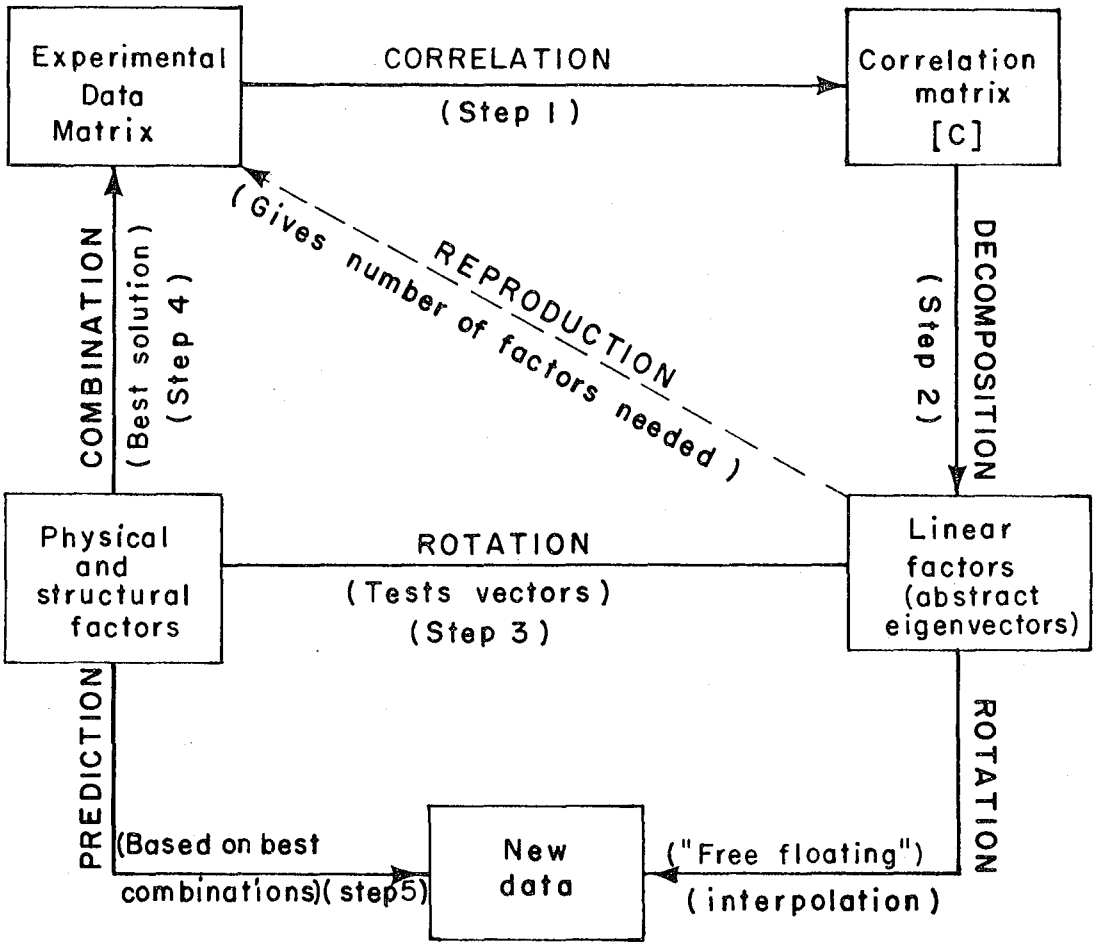
where r_k and c_k represent mutually independent row and column factors. Thus, in matrix notation the data matrix may be expressed as the product of a row and a column-related matrix:

$$[D] = [R] \cdot [C] . \quad (8)$$

The procedure used in FA may be analyzed stepwise as shown in Fig. 2 and outlined below:

1. Correlation: an experimental data matrix is used to construct a correlation matrix.
2. Decomposition: the correlation matrix is decomposed into a number of linear factors (abstract eigenvectors) capable of reproducing the data points within experimental error (reproduction).
3. Rotation: relates physically signficant parameters to the abstract factors generated in the preceding operation.

Factor analysis.



XBL779-6919

Figure 2. Factor analysis.

4. Combination: the ultimate goal of FA is to find sets of real factors (vectors) that can be combined to reproduce a data matrix within the required precision (best solution).
5. Prediction: once one has obtained good solutions, it is possible to predict new data by applying the free-floating technique for obtaining missing data points,[15,17] using key vectors and loadings obtained from combinations developed.

This technique has been used in a large number of chemical systems[16b,c] and in interactive studies of molecule-biological test pairs.[17a,b]

PATTERN RECOGNITION (PR)

The third predictive technique selected, pattern recognition (PR), is an expanding branch of artificial intelligence (AI) long familiar to engineers, biologists, and psychologists.[18] More recently, complex chemical problems involving analyses of large quantities of data are being examined with PR techniques.[19a,b]

Pattern Recognition: Methodology*

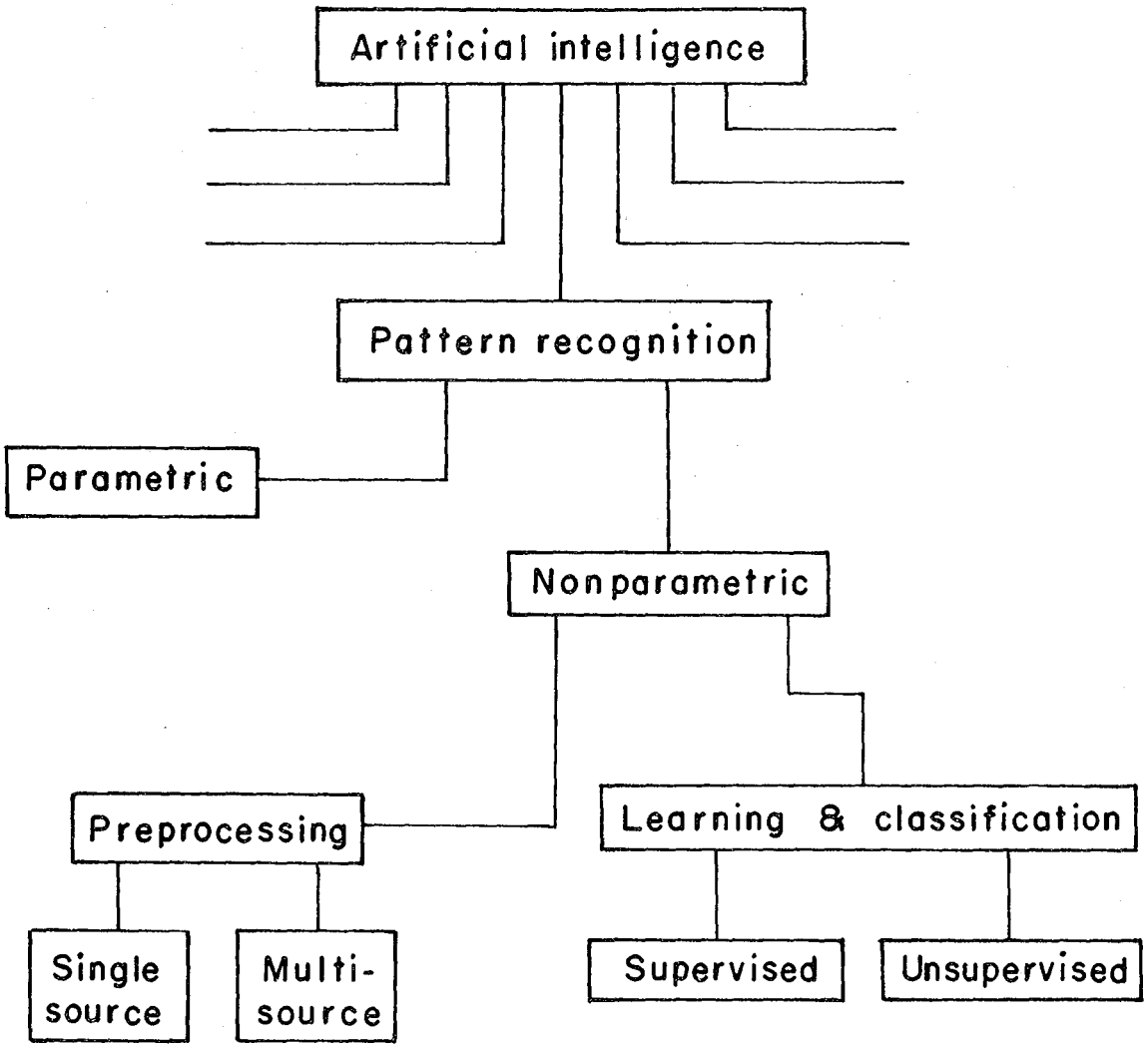
Operationally, the methods of PR fall into two classes — parametric or nonparametric (Fig. 3). Since parametric methods of PR assume access to probability density functions not usually available in practical chemical-biological interaction problems, I will confine my remarks to the nonparametric branch of PR.

First, we may consider each experimental data point in a collection as an object in n-dimensional space whose measurements are, in fact, its coordinates. Thus, the distance between any two points may be construed as a measure of their similarity. Mathematically, this similarity between objects X_i and X_j is expressed as

$$d_{ij} = \left(\sum_{k=1}^{m} (X_{ik} - X_{jk})^2 \right)^{1/2} \qquad (9)$$

and since similarity increases as d_{ij} approaches zero, a new similarity function

*See Refs. 19a,b for a more complete treatment of the subject.



XBL7710-6920

Figure 3. Functional analysis of pattern recognition techniques.

is defined, such that similarity decreases with interpoint distance:

$$S_{ij} = 1 - d_{ij}/D_{ij} \quad , \tag{10}$$

where D_{ij} represents the maximum distance between X_i and X_j. Next, the classification and learning processes operate on the n-space in one of two learning modes: supervised or unsupervised. In the former mode some of the points are identified (classified) and function as a "training set," which can then be used to classify unknown points, using a classification rule derived from the training set. In unsupervised learning there is no training set. Instead, the objective is to locate clusters of points in n-space which serve as clues to possibly significant relationships.

The preprocessing step shown in Fig. 3 entails changing the actual structure of points in n-space and will not be described here. For unsupervised learning, preprocessing is minimal, generally being confined to autoscaling of measurements with different units to obtain equal weighting, regardless of the units employed.

Mapping and Display of Data

If the parameters of a system have been judiciously selected with regard to the property being studied, like objects will be identified by their similar measurements, hence their proximity in n-space. However, for $n > 3$, computer techniques can be used to reduce the data to a more manageable two- or three-dimensional space. Here the technique of nonlinear mapping (n&m), which attempts to preserve interpoint distances in the ordered space, is useful. Figure 4 illustrates the acid-base separation achieved in a data set abstracted from the periodic table and using six properties to describe each element ($n = 6$), none of which, used alone, is capable of achieving this separation.

The pattern recognition techniques described above have been applied by Kowalski and Bender to the screening of anti-cancer drugs.[20] They should also prove useful in addressing the complex, multivariate toxicity problems encountered in chlorinated industrial effluents.

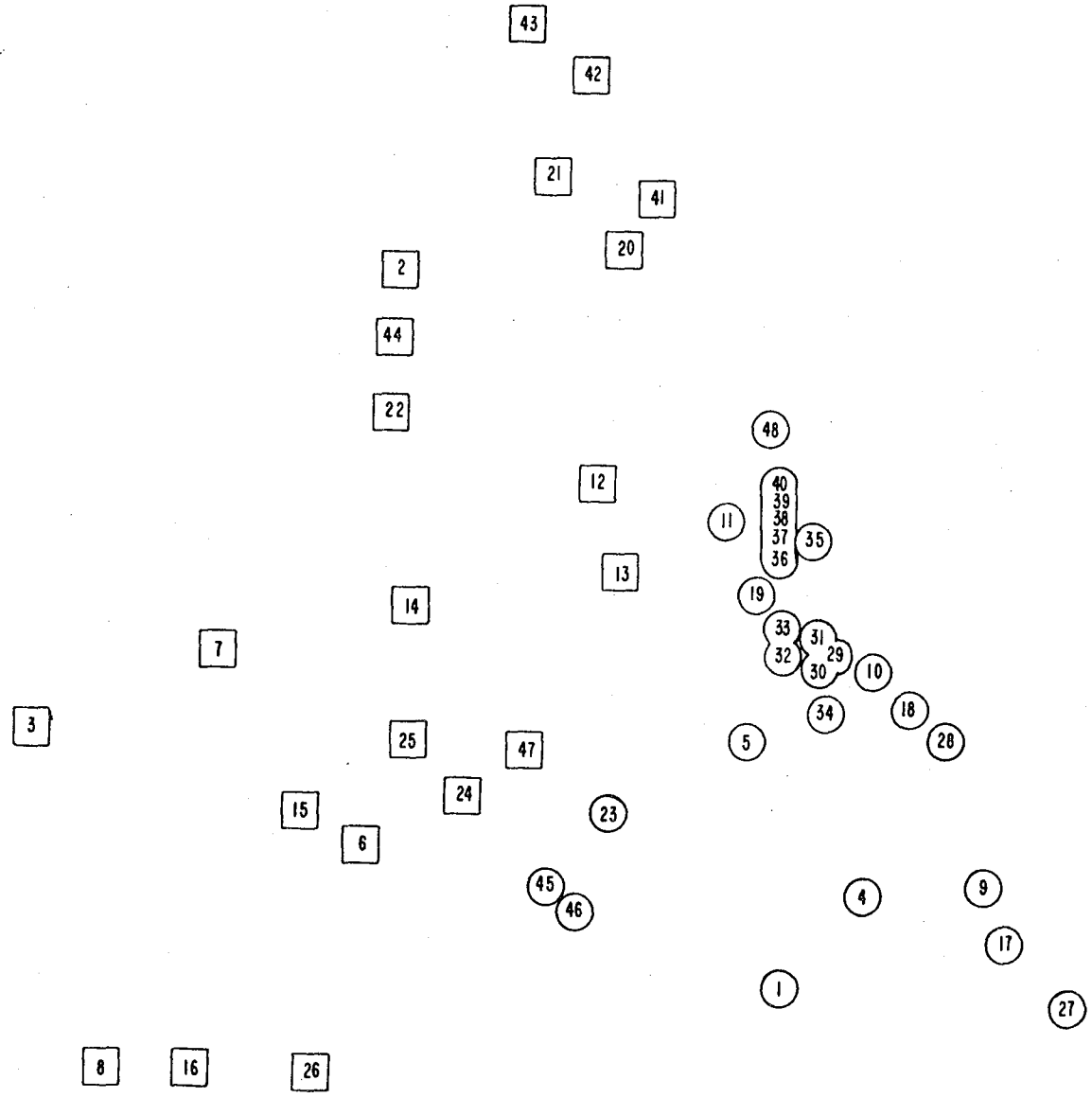


Figure 4. Acids (□) and bases (○), nln from 6-space to 2-space (Ref. 19a).

XBL7710-6917

MOLECULAR CONNECTIVITY (MC)

Of the four techniques selected for their predictive potential, molecular connectivity, which is based on topological principles, is perhaps the least familiar. The organic chemist has long used topology, since a structural formula is in reality a topological graph containing structural information related to molecular bonding (connectivity), branching, size, and shape. Chemists have also long known that even minor structural variations in molecules, such as ortho versus meta substitution in a benzene ring, can have profound effects on physical properties (melting point, boiling point), on chemical reactivity, and on biological toxicity. The question then arises: is it possible to differentiate molecular structures to such an extent, by some abstract numerical means, that correlations with physical, chemical, and biological properties become feasible? Molecular connectivity constitutes just such an attempt to evaluate molecular structure quantitatively.[21] It is defined by Kier as a "non-empirical derivation of numerical values that encode within them sufficient information to relate to many physicochemical and biological properties."[21]

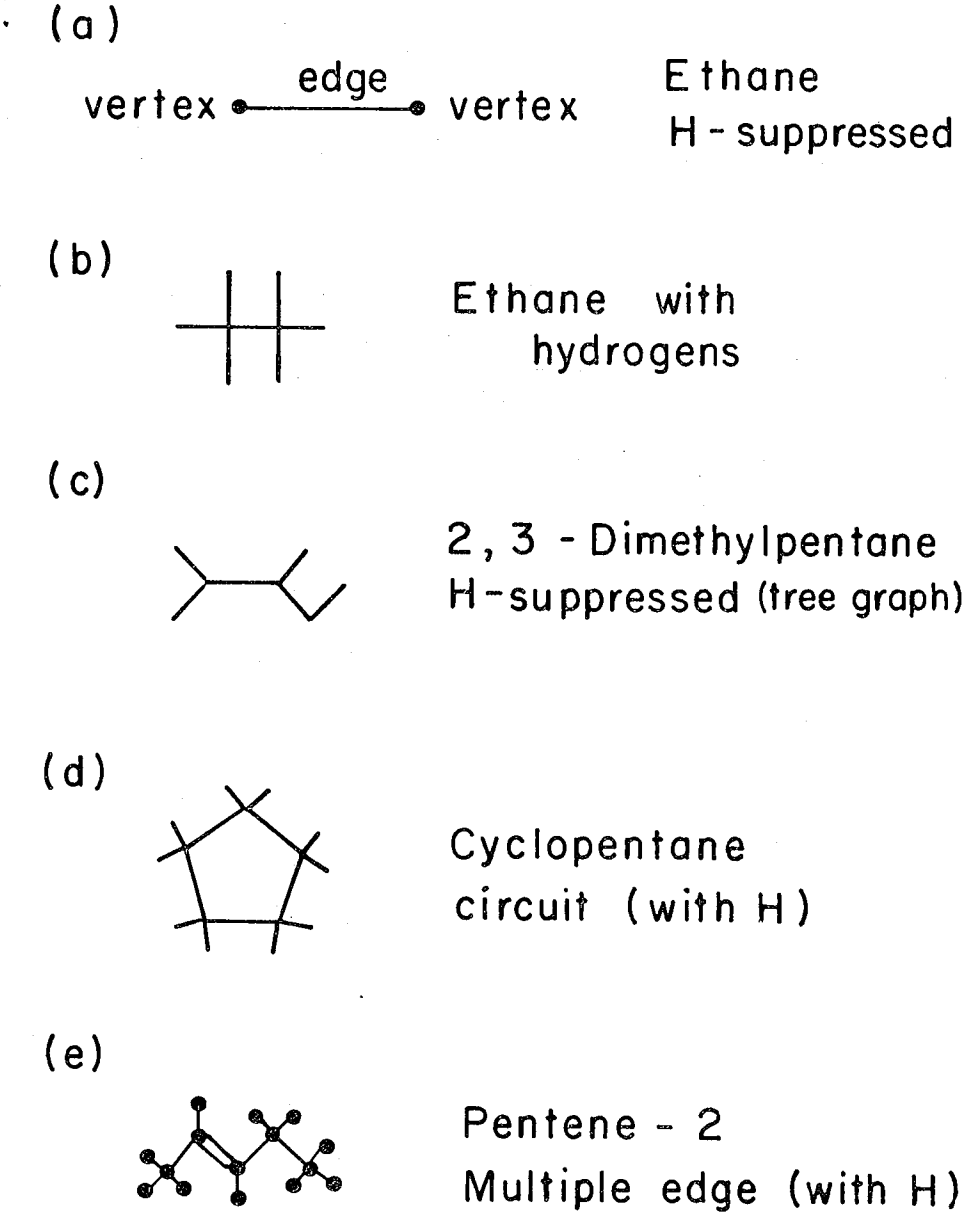
Some Definitions and Simple Graph Theory

A graph is a set of points (vertices) connected by lines (edges) (Fig. 5a). Graphical representations of some organic compounds are shown in Fig. 5. The chemical graph is a topological matrix. In its H-suppressed form, the graph, numbered in any order, may be converted to a matrix array in which vertex numbers correspond to row and column entries (T_{ij}) of 0 or 1, denoting the absence or presence of an edge between vertices i and j:

$$\delta_i = \sum_{j=1}^{n} T_{ij} \quad . \tag{11}$$

This topological matrix lends itself to mathematical operations from which numbers characteristic of the graph may be abstracted and a topological index developed.*

*See Chapter 2 of Ref. 21 for an excellent review of this subject.



XBL7710-6918

Figure 5. Graph representations of chemical structures.

The Molecular Connectivity Method

The Molecular Connectivity (MC) method assumes that information essential to a quantitative correlation of the organic molecular structure with the properties in question is inherent in a valence-weighted graph (G_v). Secondly, a relationship between the connectivity characteristics of the graph and the specified molecular properties is postulated. This relationship is expressed as a sum of terms, each linearly dependent on the graph characteristics.

The connectivity function $C(\chi)$ for a given graph may be written as

$$C(\chi) = b_o + \sum_{m,t} b_t(m)\,{}^m\chi_t \quad . \tag{12}$$

Here $b_t(m)$ depends on the property in question and may be calculated from a model or from theory, or can be obtained by multiple regression using experimental data. In the latter case, the experimental values are regressed against $C(\chi)$. The number of edges in G_v determines the highest order of the χ term. Each connectivity index term ${}^m\chi_t$ is defined by its subgraph type ($\underline{t}$) of $\underline{m}$ connected edges with subgraph order $\underline{m}$. Subgraphs are of the four types listed in Table 1.

TABLE 1. Subgraph types.

Type	Notation	Valency	Descriptor
1	t = P	≤ 2	Path
2	t = C	$3 = v = \leq 4$	Cluster (star — special case)
3	t = PC	2 + 3/4	Path/cluster
4	t = CH	--	Chain/circuit (at least one cycle)

Connectivity indices ${}^m\chi_t$ are obtained by summing terms over all distinct subgraphs.

$${}^m\chi_t = \sum_{j=1}^{n_m} {}^mS_j \quad , \tag{13}$$

where n_m is the number of type $\underline{t}$ subgraphs of order $\underline{m}$; ${}^mS_j$ terms are calculated for each subgraph, as reciprocal square root functions of valency:

$$ {}^{m}S_j = \prod_{i=1}^{m+1} (\delta_i)_j^{-\frac{1}{2}} \quad , \tag{14} $$

where j refers to a particular set of edges. The number of valences involved depends on subgraph type. Summation terms ${}^{0}\chi$ through ${}^{4}\chi$ are shown in Table 2.

TABLE 2. ${}^{m}\chi_t$ indices.

Term order, m	Vertex number, n	Path types	Equation
0	1	P	${}^{0}\chi = \sum_{i=1}^{n} \delta_i^{-\frac{1}{2}}$
1	2	P	${}^{1}\chi = \sum_{s=1}^{N_e} (\delta_i \delta_j)^{-\frac{1}{2}}$
2	3	P	${}^{2}\chi = \sum_{s=1}^{n_m} (\delta_i \delta_j \delta_k)_s^{-\frac{1}{2}}$
3	4	P, C, CH	${}^{3}\chi_t = \sum_{s=1}^{n_m} (\delta_i \delta_j \delta_k \delta_\ell)_s^{-\frac{1}{2}}$
4	5	P, C, P/C, CH	${}^{4}\chi_t = \sum_{s=1}^{n_m} (\delta_i \delta_j \delta_k \delta_\ell \delta_p)_s^{-\frac{1}{2}}$

From these it may be seen that the zero order subgraph consists of a single vertex (no edges); ${}^{1}\chi$ is summed over all edges, appropriately weighted by reciprocal square root valencies. Here we have only one type of graph edge. Second order subgraphs have pairs of adjacent edges of single path type P. Thus, each term will contain the reciprocal square root product of three vertex valencies.

In the third order connectivity index (${}^{3}\chi$), path cluster and chain terms may occur. Finally, in ${}^{4}\chi$ all four subgraph types are possible for the first time. Here n_m in the summation term shown refers to the number of type $\underline{t}$ subgraphs having four edges. ${}^{m}\chi_t$ terms of higher order are calculated in a similar manner. Figure 6 illustrates steps in the calculation of the first order connectivity index ${}^{1}\chi$ for two isomeric branched aliphatic hydrocarbons (n = 7): 2,2,3-trimethylbutane and 2,4-dimethylpentane. By way of illustration, the topological matrices and algorithm for dimethylcyclohexane and subgraphs are shown in Fig. 7.

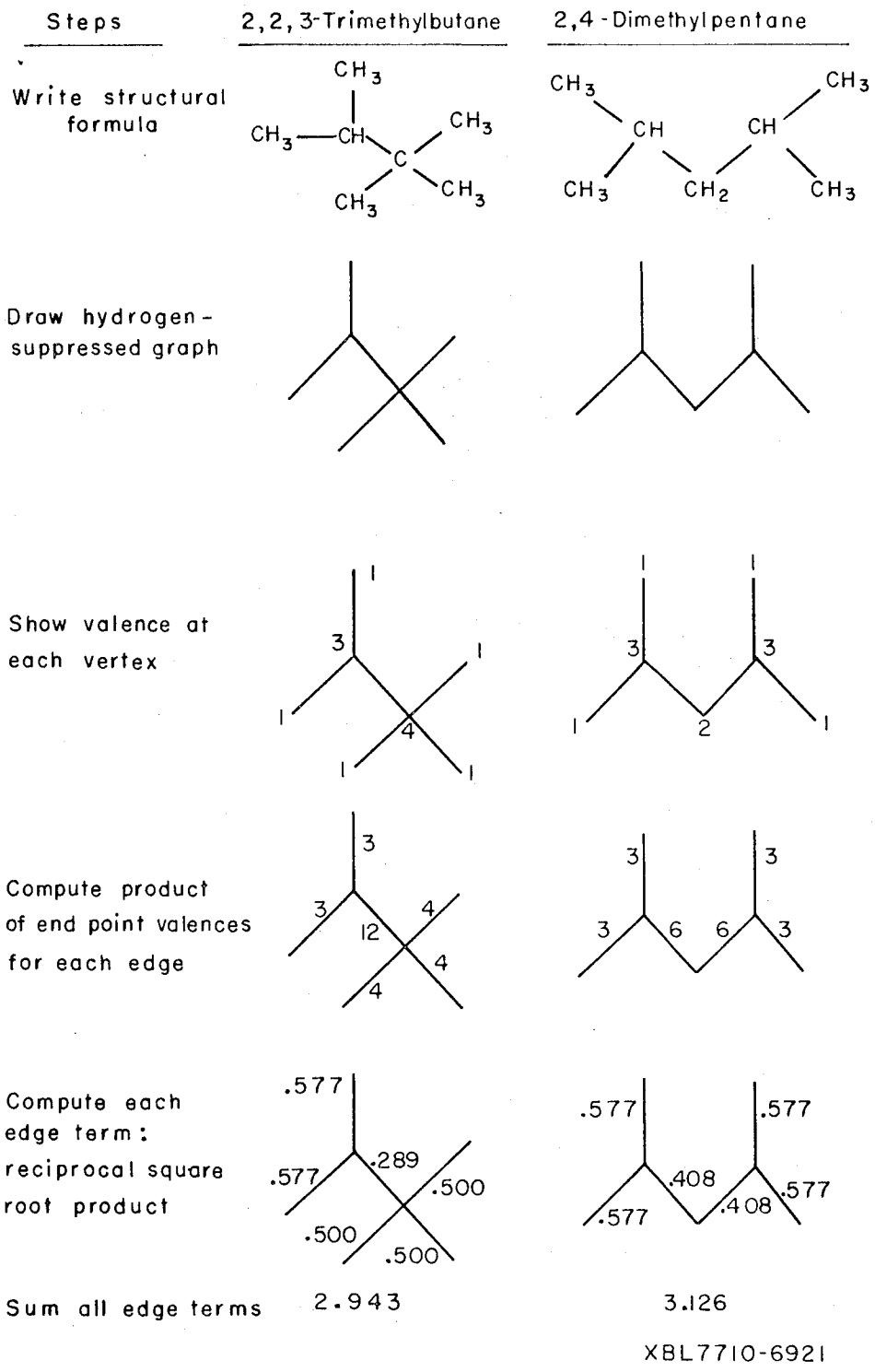
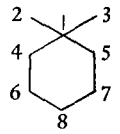
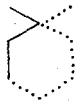
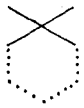
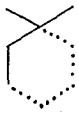
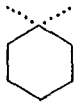


Figure 6. Procedure for calculating connectivity index ^{1}X (Ref. 21).

Example Graph	Topological Matrix
	$\begin{pmatrix} 0&1&1&1&1&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&1&0&0 \\ 1&0&0&0&0&0&1&0 \\ 0&0&0&1&0&0&0&1 \\ 0&0&0&0&1&0&0&1 \\ 0&0&0&0&0&1&1&0 \end{pmatrix}$

Vertex Number Sets	Subgraph Matrix	v_i^S	Subgraph	Type
1 2 4 6	$\begin{pmatrix} 0&1&1&0 \\ 1&0&0&0 \\ 1&0&0&1 \\ 0&0&1&0 \end{pmatrix}$	2 1 2 1 $E_S = 3$		Path
1 2 3 4 5	$\begin{pmatrix} 0&1&1&1&1 \\ 1&0&0&0&0 \\ 1&0&0&0&0 \\ 1&0&0&0&0 \\ 1&0&0&0&0 \end{pmatrix}$	4 1 1 1 1 $E_S = 4$		Cluster (star)
1 2 3 4 6	$\begin{pmatrix} 0&1&1&1&0 \\ 1&0&0&0&0 \\ 1&0&0&0&0 \\ 1&0&0&0&1 \\ 0&0&0&1&0 \end{pmatrix}$	3 1 1 2 1 $E_S = 4$		Path/Cluster
1 4 5 6 7 8	$\begin{pmatrix} 0&1&1&0&0&0 \\ 1&0&0&1&0&0 \\ 1&0&0&0&1&0 \\ 0&1&0&0&0&1 \\ 0&0&1&0&0&1 \\ 0&0&0&1&1&0 \end{pmatrix}$	2 2 2 2 2 2 $E_S = 6$		Circuit

(E_S = subgraph edge count)

XBL 7711-10514

Fig.7. Subgraph evaluation and algorithm for dimethylcyclohexane(Ref. 21).

Connectedness values were determined from edge counts E_s:

$$E_s = \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \quad , \tag{15}$$

where A is the adjacency matrix.

Some General Observations on the MC Method

Using the techniques described, Kier and others have successfully correlated MC with physical[22a-c] and biological[23a,b] properties. The method has the advantage of relative simplicity and flexibility. It can be used to represent molecular structure quantitatively at a number of levels of complexity. Each level provides some information uniquely related to the structure (graph, subgraph) and through it to physical, chemical, and biological characteristics.

The C(X) function is essentially a weighted count of substructures of the molecule, each described numerically with reference to adjacencies within them. Not discussed in this paper is the manner in which MC incorporates hetero atoms and valence differences between them. $^1\chi$ takes account only of adjacent influences on a specific atom. These are modified in the higher order subgraph terms. Also basic to the χ calculation is $\delta_i\delta_j$ — the atom product — and use of the reciprocal square roots of this product.

Preliminary attempts to develop an atomic chi value $c(^1\bar{\chi}_i)$ in which each bond term (c_{ij}) is divided equally between the two vertices and the half-bond terms summed, have further fine-tuned the method. Other structural aspects which require further refinement are: cis-trans isomerism, nonbonded steric interactions and conformational structure, all of which have either three-dimensional or directional features, or both, that are not included in the original treatment. The fact that excellent correlations have already been achieved for fairly complex systems largely within the limits of an elementary graphical approach, bodes well for the future of MC as a still maturing technique for coping with the multifactorial problems of chemical-biological interactions.

SUMMARY AND CONCLUSIONS

Four techniques for assessing the multifactorial problems of toxicity and carcinogenicity have been briefly described: structure/activity relationships (SAR), factor analysis (FA), pattern recognition (PR), and molecular connectivity (MC). While it is clear that none provides easy answers, it would appear that the more recent areas of PR and MC both merit more intensive investigation as predictive tools. In particular, the relative simplicity of the MC approach and the possibility of substantially reducing the empirical component are attractive incentives for pursuing further work in this area.

REFERENCES

1. Public Law 94-469, "Toxic Substances Control Act," 90 STAT.2003-51 (Oct. 11, 1976).
2. Barnes, J.M. and Denz, F.A., "Experimental Methods Used in the Determination of Chronic Toxicity," Pharmacol. Rev. 6(2), 191-242 (1954).
3. (a) Murray, C., "Senate Panel Probes Kepone Disaster," Chem. Eng. News, 17-18 (Feb. 2, 1976); (b) Anon., "Kepone Confirmed as Long Term Problem," ibid (Aug. 23, 1976); (c) "OSHA, EPA, FDA to Limit Worker, Public Exposure to DBCP Pesticide," EPA News Release (Thurs., Sept. 8, 1977).
4. Carter, L.J., "Michigan's PBB Incident: Chemical Mix-up Leads to Disaster," Science 192, 240-243 (April 1976).
5. Kland, M.J., "The VC-PVC Crisis: A Systematic Approach to Toxicological Problems," 30th Northwest Regional Meeting, Am. Chem. Soc., Honolulu, Hawaii, June 12-13, 1975, Lawrence Berkeley Laboratory Report LBL-3275 (August 1975).
6. (a) Crum Brown, A. and T.R. Fraser, V. "On the Connection Between Chemical Constitution and Physiological Action; Part I. On the Physiological Action of (the) Salts of (the) Ammonium Bases Derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, Nicotia," Trans. Roy. Soc. Edinburgh 25, 151-203 (1868). (b) XX Part II. "On the Physiological Action of the Ammonia Bases Derived from Atropia and Conia," ibid. 693-739 (1869).
7. Hansch, C., "On the Structure of Medicinal Chemistry," J. Med. Chem. 19(1),1-6,(1976).
8. (a) Hammett, L.P., Physical Organic Chemistry (McGraw-Hill, New York, 1940),

Chapter VII; (b) Hammett, L.P., "Effect of Structure Upon the Reactions of Organic Compounds: Benzene Derivatives," J. Am. Chem. Soc. 59, 96-103 (1937).

9. (a) Hansch, C. and T. Fujita, "ρ-σ-π Analysis. A Method for (the) Correlation of Biological Activity and Chemical Structure," J. Am. Chem. Soc. 86, 1616-26 (1964); (b) Hansch, C. and W. J. Dunn, III, "Linear Relationships Between Lipophilic Character and Biological Activity of Drugs," J. Pharm. Sci. 61(1), 1-19 (Jan. 1972).
10. (a) Metcalf, R.L., "A Century of DDT," J. Agr. & Fd. Chem. 21(4), 511-519 (July/Aug. 1973); (b) Fahmy, M.A.H., R. Fukuto, Robert L. Metcalf, and Roy L. Holmstead, "Structure-Activity Correlations in DDT Analogs," J. Agr. Fd. Chem. 21(4), 585-591 (1973).
11. Paris, D.F. and D.L. Lewis, "Chemical and Microbial Degradation of Ten Selected Pesticides in Aquatic Systems," Residue Reviews 45, 95-124; F. A. Gunther and J. D. Gunther, editors (Springer-Verlag, New York, 1973).
12. (a) Gabliks, J., E. M. Askari, and N. Yolan, "DDT and Immunological Responses. I. Serum Antibodies and Anaphylactic Shock in Guinea Pigs," Arch. Env. Health 26, 305-308 (June 1973); (b) Askari, E.M. and J. Gabliks, "DDT and Immunological Responses, II. Altered Histamine Levels and Anaphylactic Shock in Guinea Pigs," Arch. Env. Health 26, 309-312.
13. Taft, R.W., Jr., in Steric Effects in Organic Chemistry, M.S. Newman, editor (Wiley, New York, 1956), pp. 556-675.
14. Rummel, R.J., Applied Factor Analysis (Northwestern Univ. Press, Evanston, 1970).
15. Howery, D.G., "Factor Analyzing the Multifactor Data of Chemistry," American Laboratory 14-25 (February 1976).
16. (a) Weiner, P.H., E.R. Malinowski and A.R. Levinstone, "Factor Analysis of Solvent Shifts in Proton Magnetic Resonance," J. Phys. Chem. 74(26) 4537-42 (1970); (b) Weiner, P.H. and D. G. Howery, "Factor Analysis of Some Chemical and Physical Influences in Gas-Liquid Chromatography," Anal. Chem. 44(7), 1189-94 (June 1972); (c) Weiner, P.H. and J.F. Parcher, "Prediction of Some Physical Properties of Organic Molecules by Factor Analysis of Gas Chromatographic Retention Indices," Anal. Chem. 45(2), 302-307 (1973).

17. (a) Weiner, M.L. and P.H. Weiner, "A Study of Structure-Activity Relationships of a Series of Diphenylaminopropanols by Factor Analysis," J. Med. Chem. 16(6), 655-661 (1973); (b) Wooton, R., R. Cranfield, G.C. Sheppey and P.J. Goodford, "Physicochemical-Activity Relationships in Practice. 2. Rational Selection of Benzenoid Substituents," J. Med. Chem. 18(6), 607-613 (1975).

18. MacDonald, J.C., "Time-Shared Pattern Recognition," American Laboratory 31-34 (Feb. 1977).

19. (a) Kowalski, B.R., and C.F. Bender, "Pattern Recognition. A Powerful Approach to Interpreting Chemical Data," J. Am. Chem. Soc. 94(16), 5632-39 (1972); (b) "Pattern Recognition. II. Linear and Nonlinear Methods for Displaying Chemical Data," J. Am. Chem. Soc. 95, 686-693 (1973).

20. Kowalski, B.R. and C. F. Bender, "The Application of Pattern Recognition to Screening Prospective Anticancer Drugs," J. Am. Chem. Soc. 96, 916-918 (1974).

21. Kier, L.B. and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, (Academic Press, New York, 1976), Preface, p. xi.

22. (a) Hall, L.H., L. B. Kier and W.J. Murray, "Molecular Connectivity II: Relationship to Water Solubility and Boiling Point," J. Pharm. Sci. 64(12), 1974-77 (Dec. 1975); (b) Murray, W.J., L.H. Hall and L.B. Kier, "Molecular Connectivity III: Relationship to Partition Coefficients," J. Pharm. Sci. (12), 1978-81; (c) Kier, L.B., W.J. Murray, M. Randić and L.H. Hall, "Molecular Connectivity V: Connectivity Series Concept Applied to Density," J. Pharm. Sci. (8), 226-230 (August 1976).

23. (a) Kier, L.B., L.H. Hall, W.J. Murray and M. Randić, "Molecular Connectivity I: Relationship to Nonspecific Local Anesthesia," J. Pharm. Sci. 64(2), 1971-4 (Dec. 1975); (b) Kier, L.B. and W.J. Murray, "Molecular Connectivity II: Relationships to Biological Activities," J. Med. Chem. 18(12), 1272-4 (1975).

This report was done with support from the United States Energy Research and Development Administration. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the United States Energy Research and Development Administration.