Acta Chemica Scandinavica B 32 (1978) 536—542

Statistical Analysis of the Partition Coefficient

WILLIAM J. DUNN, III * and SVANTE WOLD

Research Group for Chemometrics, Umeå University, S-901 87 Umeå, Sweden

Data for 26 solutes and 6 solvents systems were analyzed by principal component analysis to get information on the partitioning process. The results indicate that the partitioning process is a two-component process. The possible origin of these components and their implications in the use of the partition coefficient in structure-activity studies are discussed.

The early work of Meyer [1] and Overton [2] and the more contemporary work of Hansch [3] have established the logarithm of the partition coefficient (log P) of biologically significant molecules as being an important parameter in determining their pharmacological properties. This parameter is determined as the equilibrium constant for a molecule, A, between an aqueous phase and a nonpolar phase. The partition

$$A_{aqueous} \rightleftharpoons A_{nonpolar}$$

$$P = \frac{[A_{nonpolar}]}{[A_{aqueous}]}$$

coefficient models the relative affinity of the molecule for lipoidal phases in biological organisms. Such nonpolar phases as diethyl ether, chloroform, 1-octanol, *etc.*, have been proposed to model these lipoidal phases with 1-octanol the most commonly used.

It was first shown by Collander [4] that for solutes of similar structure, the log P for a particular solute in one solvent, *e.g.*, chloroform, could be related to its log P values in other solvents, *e.g.*, diethyl ether, by a relationship shown as eqn. 1. This relationship has been shown to

$$\log P_{chloroform} = a + b \log P_{diethyl\ ether} \quad (1)$$

hold for a large number of solvents [5] and as such represents an example of an extrathermodynamic relationship (ETR).

An alternative formulation emphasizing the ETR nature of this relation would be eqn. 2 below. Here P_{ik} is the measured partition

$$\log P_{ik} = a_i + b_i \pi_k + e_{ik} \quad (2)$$

constant for molecule k in solvent i, a_i and b_i are solvent specific parameters and π_k is a molecule specific parameter, the so-called Hansch π-constant. The residuals e_{ik} contain the part of the measured log P_{ik} data *not* explained by the model, the random part.

This forms the basis for Hansch's definition of the π-constant with P_H the partition coef-

$$\pi_X = \log P_X - \log P_H \quad (3)$$

ficient for a parent molecule in a series such as benzene and P_X the partition coefficient for a substituted benzene such as benzoic. Hence, π_X is a substituent constant and it is assumed to be relevant only if no second-order interactions between the group and the parent nucleus occur. The π-constant and log P have been shown to model the effect of drugs on various biological systems at many levels of organization, from antigen-antibody interactions [6] to their effects on living systems.[7]

The basis of phenomenological relations such as eqns. 1 to 3 is always of interest to physical organic chemists. In the present case a better understanding of π-values could indirectly have implications about predictability of the behavior of a biological system from parallel observations made on a model of that system, *e.g.* 1-octanol.

* On leave from the University of Illinois/Medical Center, Department of Medicinal Chemistry, College of Pharmacy, 833 S. Wood, Chicago, Illinois 60680, U.S.A.

The partitioning of a substance between an aqueous and nonpolar phase is generally considered to be governed mainly by the increase in entropy due to desolvation of the solute when the substance is transfered into the nonpolar phase from water.[8,9] The low affinity for aqueous phases by relatively nonpolar solutes that results under such conditions is termed the "hydrophobic effect". On the basis of an analysis of this effect using linear regression methods, Cramer [10] has recently observed some discrepencies in this theoretical explanation of the effect.

Data related to ETR's lend themselves well to principal components analysis (PCA). Recent results in this laboratory with the use of PCA in analyzing the Hammett equation [11–14] led us to try this analysis on the partioning process hoping that this would lead to new information regarding its origin.

ETR's as Principal Components Models. It has recently been shown that ETR's of the type in eqns. 1 and 3 can be formulated as special cases of a general PC model as shown in eqn. 4.[11] In the example discussed here y refers to a

$$y_{ik} = m_i + \sum_{a=1}^{A} b_{ia}\, u_{ak} + e_{ik} \qquad (4)$$

logarithmic, measured partition coefficient in system i for solute k, m_i is the mean value of the y_i's, A is the number of components (factors) in the model, b_{ia} is a parameter associated with the variables (nonpolar phases) while u_{ak} is a parameter associated with the solutes. e_{ik} is the residual for y_{ik} and contains both experimental and model errors.

We see that eqn. 1 is formulated to imply a one component model, *i.e.*, $A = 1$ in the PC model above. However, with log P values of a number of compounds measured in a number of solvents, it is possible to *test* whether $A = 1$ or whether additional "factors" influence the variation of log P. This is part of the scope of the present article.

The PC model has a straight-forward graphical representation. (Fig. 1).

Our problem is one in 6-dimensional space which is impossible to illustrate but a 3-dimensional representation is sufficient as an analogy. In the simplest case with $A = 1$ in a PC model, the log P's of the solutes will be related by a straight line while with $A = 2$ (as is shown in the Figure) the points will lie close to a plane. The u_a's are related to the position in this plane describing the solutes. A residual standard deviation can be calculated for each object and this is the orthogonal distance of the object from the plane.

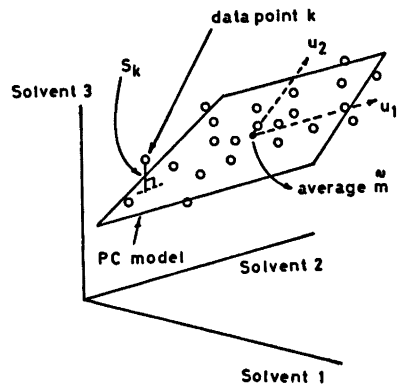

Fig. 1. Graphical representation of the Principal Component Model.

Analysis of a log P *data matrix.* For a given data matrix, the data analysis involves two main estimation problems. First, the number of significant components, the product terms in eqn. 4, A, is to be determined. Second, for the found value of A, the values of the parameters m_i, b_{ia} and u_{ak} for $i = 1, 2, ...M$; $k = 1, 2, ...N$ and $a = 1, 2, ...A$, are to be determined as to minimize the residuals e_{ik}. The details of these estimation calculations have been published elsewhere [15] and will not be presented here.

The data analyzed in this study were obtained from the Pomona College Medicinal Chemistry Data Bank.[16] They consisted of the log P values for 26 solutes in 6 aqueous/nonpolar systems. In some cases more than one value was reported for a solute in one solvent so some data selection was necessary. In those cases the median value was selected. The solvents chosen were 1-octanol, chloroform, heptane, benzene, carbon tetrachloride and diethyl ether. For 9 of these solutes log P values were not available in the solvent heptane. Heptane was included since it contains no polar functions and we were interested to see if it were significantly different from the other solvents containing polar groups

Table 1. Partition coefficient data (log P) in different solvents.

Solute	1-Octanol	Chloroform	Heptane	Carbon tetrachloride	Benzene	Diethyl ether
1. CH_3OH	−0.77 [17]	−1.26 [18]	−2.80 [18]	−2.10 [18]	−1.89 [18]	−0.85 [19]
2. C_2H_5OH	−0.37 [17]	−0.85 [18]	−2.10 [18]	−1.40 [18]	−1.62 [18]	−0.58 [19]
3. CH_3COOH	−0.31 [20]	−1.60 [21]	−2.84 [22]	−2.45 [23]	−2.00 [22]	−0.30 [19]
4. C_2H_5COOH	0.25 [20]	−0.79 [22]	−1.46 [23]	−1.79 [23]	−1.22 [24]	0.23 [25]
5. $CH_3CH_2CH_2OH$	0.25 [17]	−0.40 [18]	−1.15 [26]	−0.82 [18]	−0.70 [18]	0.28 [19]
6. $(CH_3)_2CHOH$	0.05 [27]	−0.35 [28]		−0.96 [30]	−0.96 [29]	−0.19 [31]
7. $(CH_3)_3N$	0.16 [17]	0.54 [32]		−0.09 [32]	−0.33 [33]	−0.34 [19]
8. $CH_3(CH_2)_2COOH$	0.79 [20]	−0.27 [24]		−1.02 [34]	−0.65 [24]	0.68 [35]
9. $CH_3(CH_2)_2CH_2OH$	0.88 [17]	0.45 [28]	−0.70 [18]	−0.44 [37]	−0.34 [36]	0.89 [19]
10. $(CH_3)_2CHCH_2OH$	0.65 [17]	0.34 [28]		−0.58 [30]	−0.08 [29]	0.84 [19]
11. $(C_2H_5)_2NH$	0.40 [17]	0.81 [32]		0.03 [32]	−0.05 [32]	−0.28 [19]
12. $CH_3(CH_2)_3CH_2OH$	1.56 [17]	1.05 [18]	−0.40 [18]	0.36 [38]	0.59 [29]	1.20 [17]
13. $(CH)_5N$	0.65 [39]	1.43 [40]		0.23 [42]	0.39 [41]	0.08 [19]
14. 4-$NO_2C_6H_4OH$	1.91 [43]	0.20 [44]	−2.00 [44]	−0.99 [45]	0.15 [45]	2.04 [19]
15. 3-$NO_2C_6H_4NH_2$	1.37 [43]	1.61 [25]	−0.57 [46]	0.45 [46]	1.31 [46]	1.71 [25]
16. C_6H_5OH	1.48 [47]	0.37 [48]	−0.82 [49]	−0.36 [51]	0.36 [50]	1.64 [19]
17. $C_6H_5NH_2$	0.90 [43]	1.42 [49]	−0.03 [52]	0.60 [54]	1.00 [53]	0.85 [51]
18. 4-$NH_2C_6H_4SO_2NH_2$	−0.72 [55]	−1.63 [56]		−2.52 [57]	−2.05 [57]	−0.82 [58]
19. $CH_3(CH_2)_4COOH$	1.88 [17]	1.05 [24]		0.57 [30]	0.67 [59]	1.97 [19]
20. $CH_3(CH_2)_4CH_2OH$	2.03 [17]	1.69 [18]	0.45 [18]	0.95 [18]	1.30 [18]	1.80 [17]
21. C_6H_5COOH	1.87 [43]	0.50 [60]	−0.22 [59]	−0.22 [30]	0.21 [24]	1.89 [19]
22. $C_6H_5CONH_2$	0.65 [43]	0.11 [61]		−1.54 [61]	−0.71 [61]	−0.22 [61]
23. 3-$NO_2C_6H_4OH$	2.00 [43]	0.41 [62]	−1.40 [44]	−0.58 [57]	0.42 [54]	2.20 [19]
24. 3-$NO_2C_6H_4COOH$	1.83 [43]	0.48 [24]	−1.22 [49]	0.15 [30]	0.21 [63]	1.97 [64]
25. 2-OH-C_6H_4COOH	2.26 [55]	0.50 [24]	−0.57 [65]	−0.15 [30]	0.45 [66]	2.37 [19]
26. 4-OH-C_6H_4COOH	1.58 [43]	−2.00 [49]	−1.82 [65]	−1.38 [65]	−1.07 [65]	1.42 [19]

which are commonly used in structure-activity studies. The partition coefficient data are given in Table 1.

I. The data were first regularized to give them zero mean and variance of unity. This is done to give each solvent system equal weight in the analysis. With the complete matrix of 26 solutes and 5 solvent systems PCA resulted in a model with $A=2$ components being significant. In this first calculation two objects, Nos. 26 (4-hydroxybenzoic acid) and 22 (benzamide) were indicated to be outliers and were removed from the analysis. The partition coefficient for object 26 in chloroform of −2.00 is unusually low while object 22 deviated considerably in its log P estimates in all solvents except benzene. With these two objects deleted the calculations were repeated. Again A was found to be 2, *i.e.*, a two-component model resulted. The parameters for the variables and objects are given in Fig. 2 and Table 2, respectively.

II. For the 17 solutes for which heptane partition data were available, a similar analysis also yielded a two-component model. These results are given in Table 2 and Fig. 3.

Table 2. b_i's for solvents.

Solvent	5 Solvents	6 Solvents
1-Octanol		
b_1	−0.44	−0.40
b_2	0.46	0.52
Chloroform		
b_1	−0.45	−0.42
b_2	−0.44	−0.26
Heptane		
b_1	–	−0.31
b_2	–	−0.44
Carbon tetrachloride		
b_1	−0.45	−0.42
b_2	−0.43	−0.32
Benzene		
b_1	−0.48	−0.43
b_2	−0.15	−0.47
Diethyl ether		
b_1	−0.42	−0.38
b_2	0.62	0.59

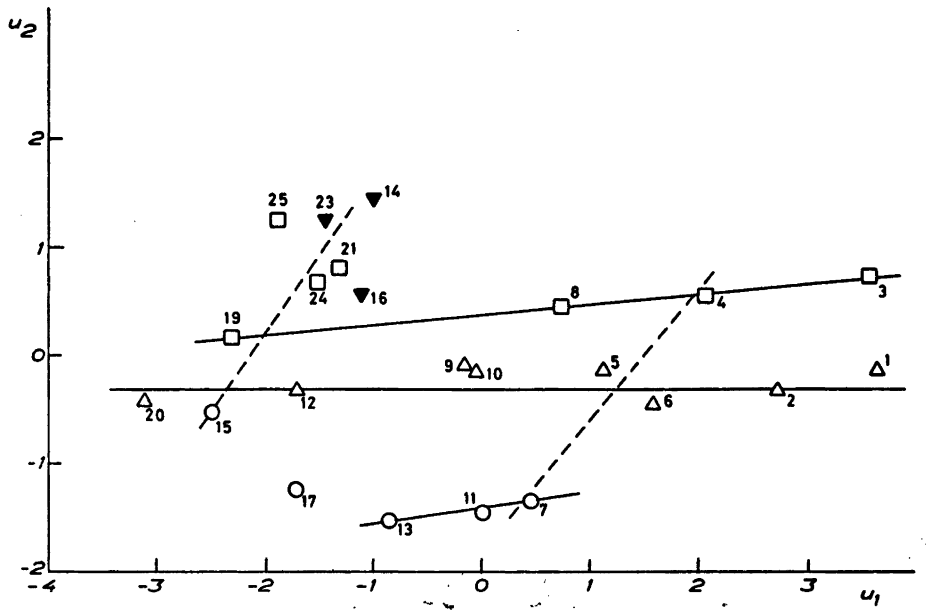

Fig. 2. Plot of the solute parameters, u_i, for PC model with 5 solvents. △ = alcohols, □ = carboxylic acids, ○ = amines, ▼ = phenols.

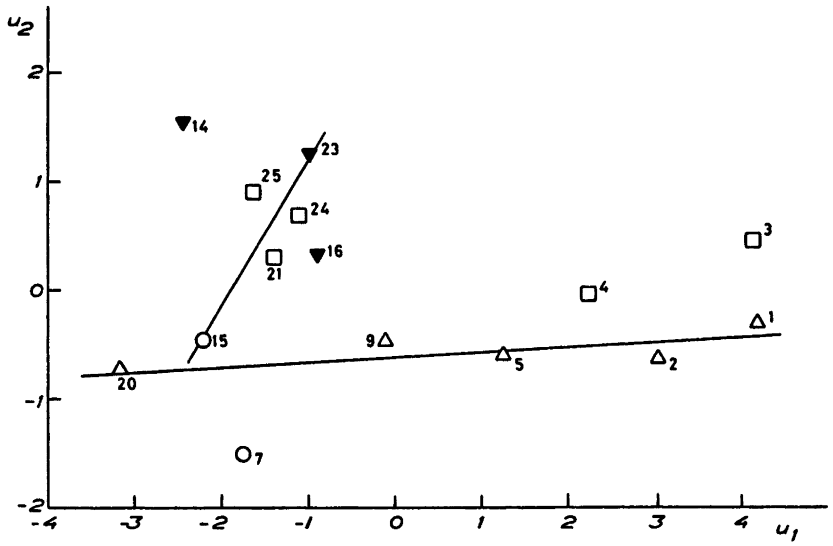

Fig. 3. Plot of the solute parameters, u_i, for PC model with 6 solvents. △ = alcohols, □ = carboxylic acids, ○ = amines, ▼ = phenols.

Graphical analysis of the solute parameters. The observation of two significant components in the partitioning process is of interest since the process is considered to be entropy controlled. By graphing u_1 *vs* u_2 for the solutes it is possible to detect regularities in these parameters that might result from structural changes within a group of solutes. The resulting plot (from analysis I) is given in Fig. 2. From Fig. 2, two observations can be made.

The largest single class of substances in the data set is the group of 8 aliphatic alcohols. These vary from $u_1 = -3.03$ for hexanol to $u_2 = 3.66$ for methanol. There is no significant variation in u_2 for this series. In the partitioning of aliphatic alcohols only a single component changes in the process with the magnitude of the component obviously related in some way to the number of carbons attached to the OH group.

An analysis of the 4 aliphatic carboxylic acids shows there is a similar relationship between carbon number and u_1. However, in this case there appears to be a small but definite decrease in the second component, u_2, with an increase in chain length. In both cases, the aliphatic alcohols and acids appear to be different from their aromatic counterparts. The partition coefficients for the phenolic objects (14, 16 and 23) and benzoic acids (objects 21, 24 and 25) contain both components. This trend is also indicated when the amines (objects 7, 11, 13, 15 and 17) are considered.

The chain length or carbon number effect associated with u_1 can most probably be attributed to the molar volume effect that is observed in partitioning experiments with homologous solutes.[10] The second component associated with u_2 is more difficult to asses but is probably the cause of the discrepencies that are observed when partition studies are extended to nonhomologous series of molecules. Its origin, mechanistically, is not possible to ascertain from this study due to limited data. However, a comparison of the aliphatic amines, alcohols and acids shows that the u_2 values increase across the series. This suggests an increase in u_2 with the electronegativity of the polar function in the series. The effect on partitioning may result from a solvent-solute interaction of a dipole-dipole type.

The structure-activity experiments of interest to the medicinal chemists are generally not ones involving homologs. The more relevant case is that in which the chemist is studying the structure-activity relationships for a series similar to solutes 15, 23 and 24. This can be considered a series of 3-substituted nitrobenzenes from which a set of π-constants can be defined.

$$\pi_X = \log P_{3\text{-X-}C_6H_4NO_2} + \log P_{C_6H_5NO_2}$$

By connecting these points in Fig. 2 we see that in such a series log P in a particular solvent will be a function of u_1 and u_2. Thus in such circumstances the partition coefficient contains two fundamental components. One component is related to the apolar nature of the solute and the other to the nature of the polar groups in the solute. This same conclusion can be made by considering the solutes 4, 5 and 7, propionic acid, propanol and trimethyl amine, respectively

Graphical analysis of the solvent parameters. The solvent parameters, b_1 and b_2, from the analysis give indications of solvent similarities or dissimilarities. On comparing the results obtained for the 5 solvents with those obtained with heptane included, it can be seen that the inclusion of heptane does not significantly alter the PC model. For the analysis with all solvents the first component is seen to be almost equally weighted with b_1 for heptane slightly less negative. There do appear to be differences in the solvents on the basis of b_2.

This can be seen by comparing b_1 and b_2 for the 6 solvents given in Table 2. On one hand 1-octanol and diethyl ether are very similar as are on the other hand chloroform, carbon tetrachloride and benzene. Heptane appears to be more like the latter group.

The similarity of 1-octanol and diethyl ether suggests that the major effect of the OH group of 1-octanol on the partitioning process is *via* a solute-solvent interaction with 1-octanol being a hydrogen acceptor.

The difference between the solvents 1-octanol and diethyl ether and the others on the basis of b_2 is very interesting. As an example we consider the three compounds 15, 23 and 24. From the derived PC model and Fig. 3 we can see that the change in y_{ik} between these compounds is related to the change in the u_i's by $\Delta y_{ik} = b_1 + 1.39\ b_2$. Comparing benzene for example, with 1-octanol, $\Delta y_{ik} = -1.08$ and $+0.32$, respectively. While most of the reduction in the standard deviation in the log P data is due to the first component in the model (59 % compared to 22 % for component 2) the second component can have a profound result in the partitioning process when different solvents are compared. For the 3 solutes compared here the partition coefficient not only changes but the change is in opposite directions. A similar result is given by the solutes 4, 5 and 7.

These results suggest that the "hydrophobic" effect observed in the solvent 1-octanol is not

comparable to one that is observed in benzene or chloroform. For structure-activity studies our results suggest two alternatives: (1) the use of one solvent only in the derivation of structure-activity correlations or, (2) the requiring of log P values for the solutes of interest in a number of solvents and use the resulting u_1 and u_2 values in the structure-activity studies. The second alternative has the advantage of allowing one to assess the relative significance of the two components in the partitioning process on the biological process being studied.

Summary. The use of principal component analysis in analyzing the partitioning process establishes that there are two fundamental components which contribute to the partition coefficient. One component is associated with the chain length in homologous series and may be the molar volume effect while the second appears to be the result of a dipolar interaction of solute and solvent. The first component appears to be the major one.

Solvents which are used to model the lipoidal phases of biological systems are of at least two different types depending on the role of the second component in partitioning. This is determining by the b_2 values derived for each solvent. It has a positive value for 1-octanol and diethyl ether and a negative value for the other solvents studied. This suggests that one should either use only one single solvent in structure-activity studies or use u_1 and u_2 values derived from several solvents.

Acknowledgement. We are grateful for support from the Swedish Natural Science Research Council and the Institute for Applied Mathematics, Stockholm.

REFERENCES

1. Meyer, H. *Arch. Exp. Pathol. Pharmakol. 42* (1899) 109.
2. Overton, E. *Vierteljahrsschr. Naturforsch. Ges. Zürich 44* (1899) 88.
3. Hansch, C. *Acc. Chem. Res. 2* (1969) 232.
4. Collander, R. *Acta Chem. Scand. 3* (1949) 717.
5. Leo, A., Hansch, C. and Elkins, D. *Chem. Rev. 71* (1971) 525.
6. Hansch, C. and Kutter, E. *J. Med. Chem. 12* (1969) 647.
7. Hansch, C. and Dunn, W. J., III. *J. Pharm. Sci. 61* (1971) 1.
8. Tanford, C. *The Hydrophobic Effect*, Wiley, New York 1973.
9. Kauzman, W. *Adv. Protein Res. 14* (1959) 1.
10. Cramer, R. D., III. *J. Am. Chem. Soc. 99* (1977) 5408.
11. Wold, S. and Sjöström, M. *Chem. Scr. 2* (1972) 49.
12. Wold, S. and Sjöström, M. *Chem. Scr. 6* (1974) 114.
13. Sjöström, M. and Wold, S. *Acta Chem. Scand. B 30* (1976) 167.
14. Sjöström, M. and Wold, S. *Chem. Scr. 9* (1976) 200.
15. Wold, S. and Sjöström, M. In Kowalski, B. R., Ed., *Chemometrics: Theory and Application*, ACS Symposium Series No. 52, 1977.
16. Leo, A. J. *Pomona College Medicinal Chemistry Project Data Bank*, Claremont, California.
17. Hansch, C. *Unpublished results.*
18. Korneman, I. and Chernorukova, Z. *Zh. Prikl. Khim. Moscow 47* (1975) 2595.
19. Collander, R. *Acta Chem. Scand. 3* (1945) 717.
20. Collander, R. *Acta Chem. Scand. 5* (1951) 774.
21. Komar, N. and Manzhelii, L. *Chem. Abstr. 65* (1966) 6523.
22. Davies, M., Jones, P., Patniak, D. and Moelwyn-Hughes, E. *J. Chem. Soc.* (1951) 1249.
23. Bartz, Q. *J. Biol. Chem. 172* (1948) 445.
24. Smith, H. and White, T. *J. Phys. Chem. 33* (1929) 1953.
25. Sandell, K. *Naturwissenschaften 53* (1966) 330.
26. Weisbordt, N., Kienzle, M. and Cooke, A. *Proc. Soc. Exp. Biol. Med. 142* (1973) 450.
27. Killingham, E., Mast, R., Bass, G. and Autian, J. *J. Pharm. Sci. 62* (1973) 22.
28. Sandell, K. *Naturwissenschaften 51* (1964) 336.
29. Tanaka, M. and Kojima, I. *J. Inorg. Nucl. Chem. 29* (1967) 1769.
30. Kuznetsova, E. and Guraii, L. *J. Phys. Chem. USSR 45* (1971) 1761.
31. Collander, R. *Plant Physiol. 7* (1954) 420.
32. Sandell, K. *Naturwissenschaften 49* (1962) 12.
33. Herz, W. and Stanner, E. *Z. Phys. Chem. 128* (1927) 399.
34. Kallossowsky, N. and Bekturov, A. *Bull. Soc. Chim. 2* (1953) 460.
35. Dermer, O. and Dermer, V. *J. Am. Chem. Soc. 65* (1943) 1653.
36. Meyer, K. and Hemmi, H. *Biochem. Z. 277* (1935) 39.
37. Meeusen, E. and Kuyskens, P. *J. Chim. Phys. 63* (1966) 845.
38. Nakano, M. and Patel, N. *J. Pharm. Sci. 59* (1970) 77.
39. Iwasa, J., Fujita, T. and Hansch, C. *J. Med. Chem. 8* (1965) 150.
40. Golumbic, G. and Orchin, M. *J. Am. Chem. Soc. 72* (1950) 4145.

41. Woodman, R. and Corbet, A. *J. Chem. Soc.* (1925) 2461.
42. Johnson, J. and Affsprung, H. *J. Phys. Chem. 72* (1968) 3223.
43. Fujita, T., Iwasa, J. and Hansch, C. *J. Am. Chem. Soc. 86* (1964) 5175.
44. Korneman, Y., Katelyanskaya, E. and Nefedova, T. *Zh. Prikl. Khim. 49* (1976) 1112.
45. Kemula, W., Buckowski, H. and Peparek, J. *Bull. Acad. Polon. Sci. 12* (1964) 347.
46. Kemula, W., Buckowski, H. and Pawlowski, W. *Bull. Acad. Polon. Sci. 12* (1964) 491.
47. Hachleid, H., Roth, S. and Seeman, P. *Biochim. Biophys. Acta 255* (1972) 178.
48. Schanker, L., Johnson, J. and Jeffery, J. *Am. J. Physiol. 207* (1964) 503.
49. Hogben, C., Tocco, D. and Brodie, B. *J. Pharmacol. Exp. Ther. 125* (1959) 275.
50. Endo, K. *Bull. Chem. Soc. Jpn. 1* (1926) 25.
51. Vaubel, J. *Prakt. Chem. 67* (1903) 477.
52. DeLugny, C., Kreutzer, H. and Visserman, G. *Recl. Trav. Chim. Pays-Bas 85* (1966) 5.
53. Farmer, R. and Wurth, F. *J. Chem. Soc. 85* (1903) 1713.
54. Kemula, W., Buchowski, H. and Pawlowski, W. *Rocz. Chem. 43* (1969) 1555.
55. Hansch, C. and Anderson, S. *J. Org. Chem. 32* (1967) 2583.
56. Sandell, K. *Naturwissenschaften 49* (1962) 12.
57. Koizumi, T., Arita, T. and Kakemi, K., *Chem. Pharm. Bull. Jpn. 12* (1964) 413.
58. Sanders, E. and Maren, T. *Mol. Pharmacol. 3* (1967) 204.
59. Brown, F. and Burg, C. *J. Chem. Soc. 123* (1923) 2430.
60. Tilly, A. and Modin, R. *Acta Pharm. Suec. 5* (1968) 311.
61. Sandell, K. *Monatsh. Chem. 89* (1958) 36.
62. Kemula, W. and Buchonski, H. *J. Phys. Chem. 63* (1959) 155.
63. *Chem. Abstr. 9* (1915) 2014.
64. Smith, H. *J. Phys. Chem. 25* (1921) 204, 605.
65. Korneman, Y. and Koroleva, T. *Zh. Prikl. Khim. 48* (1975) 1413.
66. Hendrixson, W. *Z. Anorg. Chem. 13* (1897) 73.

Received April 14, 1978.