Tetrahedron Vol. 35, pp. 1953 to 1956
Pergamon Press Ltd., 1979. Printed in Great Britain

CAUTIONS REGARDING THE PHYSICAL INTERPRETATION OF STATISTICALLY BASED SEPARATION OF THE ORTHO SUBSTITUENT EFFECT INTO INDUCTIVE, MESOMERIC, AND STERIC COMPONENTS—I

THE COMPOSITION OF TAFT E_s° CONSTANTS

H. MAGER,* P. P. MAGER and A. BARTH

Academy of Sciences of the GDR, Institute of Plant Biochemistry, DDR-401 Halle/Weinberg, Martin-Luther-University of Halle, Section of Pharmacy, DDR-401 Halle/Weinberg, and Martin-Luther-University of Halle, Section of Biosciences, Department of Biochemistry, DDR-401 Halle, Domplatz 1, DDR

(*Received in Germany* 31 *January* 1979)

Abstract—It is shown that the interpretation of the Taft E_s° values in terms of predominating inductive, mesomeric, and steric effects depends strongly upon the sample size. Nevertheless, the results obtained indicate that the E_s° constants are a function of the steric effects and to a lesser extent of the mesomeric effects of the substituents.

Within the past 20 years, multiple regression analysis has been exploited as a powerful statistical tool for the evaluation of structure-reactivity data. One goal of these analyses is the derivation of a regression equation which will provide estimates of some reaction constants for additional structure-entities within a series. A second goal is the interpretation of parameters (like σ_p, σ^+, E_s, E_s°, etc.) and reaction mechanisms in terms of predominating inductive, mesomeric, and steric effects. Recently, we have shown[1] that ordinary regression coefficients depend on the scaling of the explanatory variables (regressors) and proposed the use of standardized regression coefficients in order to get a better insight into the mechanistic features of the reaction under investigation. In this procedure the regressors are ranked in order of the sizes of these coefficients. Since the method is only straightforward if the coefficients of internal determination B_{Ij} ($j = 1, 2, \ldots, q$; q being the number of regressors. B_{Ij} is obtained when x_j is regressed on the remaining $q-1$ regressor variables.) are all alike we prefer in this paper the use of the t-values to judge the relative performances of the explanatory variables. We intend to show that the interpretation of the Taft E_s° constants depends strongly upon the nature of the substituents included in the analysis and that the conclusions drawn by Charton[2] regarding the composition of these constants seem to be not correct. E_s° constants have been defined by Taft using the hydrolysis of ortho-substituted benzoic esters.[3] In 1969, Charton[2] has reexamined E_s for aliphatic and aromatic systems. He pointed out that aliphatic E_s values are linear functions of the van der Waals radii whereas E_s° could be expressed in terms of inductive (σ_I) and mesomeric (σ_R) substituent constants.

In these analyses the σ_I constants required have been taken from the compilation of Charton[4] and the σ_R constants were obtained from the equation

$$\sigma_p = \sigma_R + \sigma_I.$$

It should be emphasized however, that often particular ortho-substituents, such as H, NO_2 and Ph have been deleted from the analyses. Since Charton did not list the values of the σ_I and σ_R constants we were forced to collect them from the literature. So the σ_I values have been taken from Charton[4] and the σ_p values from the excellent tabulation of McDaniel and Brown.[5] The values we used in our computations are summarized in Table 1.

Before presenting our results we should emphasize vigorously that a statistically significant regression equation does by no means express that there is a causal or functional relationship between the regressand and the regressor (for the terminology see Mager[6]). The only thing we can state is that in such cases the regressand approximately may be replaced by the regressors. An example for this statement is the work of Koppel[7] who was able to show that for twenty primary, secondary, and tertiary alkyl groups equation

$$E_s = 0.88 + 27.78\sigma^* - 1.90(n-3)$$

holds (with n being the number of α-C–H bonds). Without further information we may only conclude from this correlation that there is some parallelity between inductive and steric effects of alkyl groups.

In order to evaluate the relative contributions of inductive, mesomeric, and steric effects to the total ortho-substituent effect we applied the eqn (1) as proposed by Charton,[2] e.g.

$$E_s^\circ = b_0 + b_1\sigma_I + b_2\sigma_R + b_3 r_v. \qquad (1)$$

Using r_v we presumed that r_v could really be regarded as a measure of the intramolecular steric effect. It is generally assumed that the r_v values of OMe and OEt are alike and agree with that for O, an assumption which is somewhat questionable. Nevertheless, we accepted also this assumption. Charton[2] excluded the phenyl and nitro groups from the correlation for lack of a suitable van der Waals parameter. In this case the following regression equation is obtained:

$$\begin{gathered} E_s^\circ = -1.196 - 0.680\sigma_I - 3{,}088\sigma_R + 0.469r_v \\ R^2 = 0.9786 \quad s_{y.x} = 0.0965 \quad \hat{F} = 45.81 > F_{3;3;0.95} = 9.28 \\ \hat{t}(b_1) = 2.80 \quad \hat{t}(b_2) = 5.52 \quad \hat{t}(b_3) = 0.91 \quad t_{3;0.95} = 3.18 \\ N = 7. \end{gathered} \qquad (2)$$

Table 1. Substituent constants used in the correlations

Group	E_s^o	σ_I	σ_R	σ_p	r_v
OCH_3	0,99	0,25	−0,52	−0,27	1,52
$O\ C_2H_5$	0,90	0,27	−0,51	−0,24	1,52
F	0,49	0,52	−0,46	0,06	1,47
Cl	0,18	0,47	−0,24	0,23	1,75
Br	0,00	0,45	−0,22	0,23	1,85
J	−0,20	0,39	−0,12	0,27[a]	1,98
CH_3	0,00	−0,05	−0,12	−0,17	1,715[b]
NO_2	−0,75	0,76	0,02	0,78	2,18[c]/2,59[d]
C_6H_5	−0,90	0,10	−0,11	−0,01	2,535[c]/3,3[d]

[a] taken from R. A. Robinson and K. P. Ang, J. Chem. Soc. 2314 (1959).

[b] $r_{v,\min}$. [c] average van der Waals radius ($\bar{r}_v$) [d] $r_{v,\max}$.

In eqn (2) R^2 is the squared multiple correlation coefficient, $s_{y.x}$ is the residual standard deviation, $\hat{F}$ is the numerical value of the F test for significance of regression, F is the tabulated value of the F distribution with q and $N-q-1$ degrees of freedom, $\hat{t}$ is the numerical value of the t test for significance of a regression coefficient, t is the value of the t distribution with $N-q-1$ degrees of freedom, and N is the sample size. The whole analysis was performed using $\alpha = 0.05$ (type I error = probability for the rejection of a true null hypothesis). Applying the backward elimination procedure to eqn (2) gives

$$\begin{gathered} E_s^\circ = -0.298 - 0.562\sigma_I - 2.621\sigma_R \\ R^2 = 0.9727 \quad s_{y.x} = 0.0944 \quad \hat{F} = 71.32 > F_{2;4;0.95} = 6.94 \\ \hat{t}(b_1) = 2.79 \quad \hat{t}(b_2) = 11.93 \quad t_{4;0.95} = 2.78. \end{gathered} \tag{3}$$

Since in eqn (3) both the squared multiple correlation coefficient and the regression coefficients are significant at a level of 5% (cave: the level of significance has to be determined before the beginning of the analysis), eqn (3) is the final subset. It shows that for the given set of substituents E_s° may be accounted for in terms of inductive and mesomeric effects. But the interpretation of eqn (3) is rendered more difficult if the coefficients of internal determination of eqn (2) are regarded. We obtained

$$B_{I1} = 0.3087 \quad B_{I2} = 0.8450 \quad B_{I3} = 0.8401.$$

Thus σ_I and σ_R do not only explain about 97% of the variability of E_s° but also 84% of the variability of r_v (correlation coefficient $R = 0.9166$). This means that r_v may be approximately replaced by σ_I and σ_R. Since correlation coefficients of this size have been regarded as being sufficiently large to "explain" the behaviour of aliphatic E_s values, the interpretation of eqn (3) is not unambiguous. For this reason we decided to vary the sample size in order to get a statistical and physical more meaningful insight in the composition of E_s° values.

Omitting Me gives

$$\begin{gathered} E_s^\circ = 5.639 - 2.240\sigma_I + 0.8191\sigma_R - 2.430r_v \\ R^2 = 0.9918 \quad s_{y.x} = 0.0694 \quad \hat{F} = 80.55 > F_{3;2;0.95} = 19.16 \\ \hat{t}(b_1) = 2.73 \quad \hat{t}(b_2) = 0.40 \quad \hat{t}(b_3) = 1.58 \quad t_{2;0.95} = 4.30 \\ B_{I1} = 0.8823 \quad B_{I2} = 0.9922 \quad B_{I3} = 0.9906. \end{gathered} \tag{4}$$

The comparison of eqn (4) with eqn (2) shows that the mesomeric and steric contributions have changed their signs, e.g. the corresponding physical effects seem to be reversed. Although the whole equation is significant this is not true for any regression coefficient. This is an impressive example that the situation of high multicollinearity implies for the analyst that he is asking more than his data are able to answer. Backward elimination gives

$$\begin{gathered} E_s^\circ = 4.209 - 1.934\sigma_I - 1.819r_v \\ R^2 = 0.9911 \quad s_{y.x} = 0.0589 \quad \hat{F} = 167.68 > F_{2;3;0.95} = 9.55 \\ \hat{t}(b_1) = 7.82 \quad \hat{t}(b_3) = 13.95 \quad \hat{t}_{3;0.95} = 3.18. \end{gathered} \tag{5}$$

Regarding the value for B_{I3} in eqn (4) we see that 99% of the variability of r_v is explained by the inductive and mesomeric constants. Under such circumstances eqn (5) is really a product of chance. In order to obtain more precise information we decided to enlarge the sample size. Including all substituents of Table 1 in our correlations and using for $r_v(NO_2)$ and $r_v(Ph)$ the average van der Waals radius $\bar{r}_v$ (see also Kutter and Hansch[6]) we have

$$\begin{gathered} E_s^\circ = 1.61 - 0.170\sigma_I - 1.602\sigma_R - 1.023r_v \\ R^2 = 0.9656 \quad s_{y.x} = 0.1533 \quad \hat{F} = 46.85 > F_{3;5;0.95} = 5.41 \\ \hat{t}(b_1) = 0.74 \quad \hat{t}(b_2) = 3.52 \quad \hat{t}(b_3) = 4.01 \quad t_{5;0.95} = 2.57 \\ B_{I1} = 0.0334 \quad B_{I2} = 0.6370 \quad B_{I3} = 0.6318. \end{gathered} \tag{6}$$

In eqn (6) the numerical value of B_{I3} is much lower than in eqns (2) and (4). For this reason a physical interpretation is facilitated. Backward elimination leads to

$$E_s^0 = 1.486 - 1.663\sigma_R - 0.996r_v$$
$$R^2 = 0.9619 \quad s_{y.x} = 0.1474 \quad \hat{F} = 75.66 > F_{2;6;0.95} = 5.14$$
$$\hat{t}(b_2) = 3.86 \quad \hat{t}(b_3) = 4.11 \quad t_{6;0.95} = 2.45. \tag{7}$$

Equation (7) indicates that E_s^0 should be a function of both the resonance and the steric effects of the substituents. In spite of the high multicollinearities involved in eqn (2) we may regard the positive sign of b_3 as further support for the assumption that eqn (6) is more reliable. The same assertion holds then for eqn (7) when compared with eqn (3). This statement is in clear contradiction to that made by Charton,[2] who concluded that E_s^0 does not in any way represent a steric effect. Elimination of σ_R (although significant) gives

$$E_s^0 = 3.266 - 1.736r_v$$
$$R^2 = 0.8671 \quad s_{y.x} = 0.2548 \quad \hat{F} = 45.68 > F_{1;7;0.95} = 5.59.$$

Thus nearly 87% of the variability of E_s^0 may be accounted for by r_v. If the nitro and the phenyl groups are coplanar with the benzene ring to which they are attached, the $r_{v,max}$ value is the required van der Waals radius. In such a case the maximization of resonance interaction supplies a driving force for coplanarity although generally groups will preferentially take positions which will minimise the degree of steric interaction. The application of $r_{v,max}$ for the nitro and the phenyl groups confirms eqn (7), e.g. that E_s^0 may be approximately replaced by a linear combination of σ_R and r_v. However, it should be emphasized that the sample size is still too small to derive to ultimate conclusions. Besides that the aptitude of r_v as a measure for the intramolecular steric effect has not been demonstrated up to now.

Since the r_v values for OMe and OEt have been regarded as a possible source of error we omitted these groups from the correlation. The corresponding regression equation is (using $\bar{r}_v$ for NO_2 and Ph)

$$E_s^0 = 1.845 - 0.099\sigma_I - 0.905\sigma_R - 1.110r_v$$
$$R^2 = 0.9737 \quad s_{y.x} = 0.1142 \quad \hat{F} = 37.01 > F_{3;3;0.95} = 9.28$$
$$\hat{t}(b_1) = 0.57 \quad \hat{t}(b_2) = 2.04 \quad \hat{t}(b_3) = 5.75 \quad t_{3;0.95} = 3.18$$
$$B_{11} = 0.015 \quad B_{12} = 0.5146 \quad B_{13} = 0.5194.$$

Elimination of the variable with the smallest nonsignificant $\hat{t}$ value gives

$$E_s^0 = 1.784 - 0.912\sigma_R - 1.099r_v$$
$$R^2 = 0.9709 \quad s_{y.x} = 0.104 \quad \hat{F} = 66.63 > F_{2;4;0.95} = 6.94$$
$$\hat{t}(b_2) = 2.25 \quad \hat{t}(b_3) = 6.28 \quad t_{4;0.95} = 2.78.$$

After backward elimination we have

$$E_s^0 = 2.490 - 1.381r_v$$
$$R^2 = 0.9339 \quad s_{y.x} = 0.1401 \quad \hat{F} = 70.69 > F_{1;5;0.95} = 6.61.$$

In the present case E_s^0 correlates significantly with r_v alone. After replacement of $\bar{r}_v(NO_2)$ and $\bar{r}_v(Ph)$ by $r_{v,max}(NO_2)$ and $r_{v,max}(Ph)$ we obtained the following subset

$$E_s^0 = 0.771 - 1.264\sigma_R - 0.556r_v$$
$$R^2 = 0.9907 \quad s_{y.x} = 0.0587 \quad \hat{F} = 213.86 > F_{2;4;0.95} = 6.94$$
$$\hat{t}(b_1) = 6.19 \quad \hat{t}(b_3) = 11.51 \quad t_{4;0.95} = 2.78.$$

The most important results of the correlation analyses have been summarized in Table 2.

By inspection of Table 2 the following conclusions may be drawn.

(i) most of the results provide examples which indicate that E_s^0 is not appreciably affected by inductive effects (σ_I).

(ii) quantitative evidence is provided that E_s^0 are approximately a linear combination of σ_R and r_v.

(iii) in nearly all cases the simple correlation coefficient (R) between E_s^0 and r_v is greater or at least equal 0.90. Thus E_s^0 may indeed be regarded as a measure of some steric effect although some influence of the mesomeric properties of the substituents seems to be present.

Our interpretation of the results does not support the conclusions of Charton[2] that E_s^0 is only dependent upon inductive and mesomeric effects and independent of the size of the substituents.

Table 2. Composition of the best subsets obtained and simple correlation coefficients between E_s^0 and the most important parameter of the subset

Omitted Substituents	significant regressors in the subset	R^2	most significant regressor	R
NO_2, Ph	σ_I, σ_R	0,9786	σ_R	0,9998
NO_2, Ph, Me	σ_I, r_v	0,9911	r_v	0,9003
none[a]	σ_R, r_v	0,9619	r_v	0,9312
none[b]	σ_R, r_v	0,9694	σ_R	0,9245 (0,88)[*]
OMe, OEt[a]	r_v	0,9339	r_v	0,9667
OMe, OEt[b]	σ_R, r_v	0,9907	r_v	0,9498

[a] average van der Waals radius used for NO_2 and Ph.

[b] $r_{v,max}$ used for NO_2 and Ph.

[*] R for the correlation between E_s^0 and r_v.

REFERENCES

[1] H. Mager and A. Barth, *Pharmazie* in print.

[2] M. Charton, *J. Am. Chem. Soc.* 91, 615 (1969).

[3] R. W. Taft, *Separation of Polar, Steric, and Resonance Effects in Reactivity*. In *Steric Effects in Organic Chemistry* (Edited by M. S. Newman) p. 556. Wiley, New York (1956).

[4] M. Charton, *J. Org. Chem.* 29, 1222 (1964).

[5] D. H. McDaniel and H. C. Brown, *Ibid.* 23, 420 (1958).

[6] H. Mager, *Sci. Pharm.* 45, 68 (1977).

[7] I. A. Koppel, *Reakts. Sposobnost Org. Soedin.* 2, 26 (1965).

[8] E. Kutter and C. Hansch, *J. Med. Chem.* 12, 647 (1969).