

A PRIORI REFITTING OF THE TAFT EQUATION BY NON-LINEAR LEAST SQUARES METHODS*

S. Ehrenson

Chemistry Department, Brookhaven National Laboratory, Upton, New York

(Received 11 November 1963; in revised form 23 December 1963)

Taft's linear free energy equation for aromatic reactivities (1) is currently undergoing statistical reexamination to test, and, if possible, to extend, its generality. We here discuss results for the σ^0 reactivity series and delineate future efforts.

Several correlational and conceptual uncertainties arise upon separation of inductive and resonance effects by the dual substituent formulation

$$\log k/k_0 = I + R = \rho_I(\sigma_I + \bar{\sigma}_R) \quad (1)$$

For examples, the position-dependent resonance terms, R, cannot be decomposed in this manner (i.e., $\bar{\sigma}_R = R/\rho_I$) to yield a single satisfactory set of σ_R values for all types of reactivities. Nor is there any compelling reason why the inductive term, I, should be position independent (2).

To examine these and other points and their interrelations, a more general form of the Taft equation is employed

$$\log k/k_0 = \rho_I\sigma_I + \rho_R\sigma_R \quad (2)$$

The rhos are now position dependent variables, while the sigmas are not. All four parameters are assumed to be mutually independent in the general case; under obvious constraints, however, equation 2 may be transformed to equation 1. The experimental data, $\log k/k_0$ values, are processed by two complementary non-linear least squares techniques programmed for the IBM

*Work performed under the auspices of the U. S. Atomic Energy Commission.

7090 computer. The first, used primarily in unconstrained fittings, equation 2, and in validity tests for specific constraints (e.g., $\rho_I^P = C^T \rho_I^M$), involves an iterative technique to achieve parameter self-consistency, and requires initial specification only of the sigmas. The second is an adaptation of the well-known "Gauss-Newton" method, as modified by Hartley (3); it is used to refine parameter values obtained by the first procedure when particular constraints have been chosen (4). Both procedures are capable of simultaneous optimization of all parameters.

The σ^0 reaction series was chosen for initial study for several reasons. They are well fitted as a group by the Taft equation, have a fairly constant set of α values ($= R^M/R^P \sim 0.5$) and are substantially independent, as far as being fitted by the σ_R^0 scale, of the nature of the reaction, presumably all because direct resonance interactions between substituents and reaction centers are precluded. As well, their reactivity parameters correlate well with n.m.r. shielding parameters for substituted fluorobenzenes (5).

Table I displays the results of these and previous correlations on 12 σ^0 reactivity series with the 5 solvent-independent substituents, H, OCH_3 , CH_3 , F and Cl and Br (the latter pair assumed the same for present statistical purposes; experimental values were averaged). The NO_2 group was also considered for reactions in hydroxylic solvents. A number of significant points may be detailed.

1) The least-squares equation 1 results agree fairly well with those of Taft, the latter obtained essentially reaction-by-reaction. The A13 and B6 rhos differ by 10 per cent which is the largest discrepancy noted. The standard deviation, 0.0395, corresponds to an uncertainty, per measured point, of ~ 5 per cent (6), perhaps a little more than intuition would suggest for experimental uncertainty.

2) Equation 2, unconstrained, yields considerably better SD (standard

TABLE I
Dual Substituent Equation Parameters and Measures of Fit for σ^0 Reactivities

Reactions ^f	Taft Parameters		Eq. 1, LSM ^h		Equation 2, Least Squares Method ^g									
	ρ_I	α	ρ	α	No constraints			(Best σ^0, α) ^k						
					ρ_I	σ^0	α	ρ_I	σ^0	α	ρ_I	σ^0	α	
A13	2.24	0.38	2.02		2.00	0.92	0.51	0.55	2.14	0.47	0.56	2.16	0.07	0.52
B1	0.46	.50	0.48		0.48	0.98	.36	.38	0.49	.10	.52	0.50	.03	.50
B3	.24	.60	.23		.22	.99	.61	.61	.23	.07	.59	.23	.05	.56
B6	1.19	.40	1.08		1.02	.94	.53	.51	1.07	.25	.59	1.08	.04	.58
B14	0.73	.60	0.69		0.71	1.21	1.14	.84	0.68	.28	.65	0.69	.04	.64
B21	~ 1.0	>.75	1.06		1.00	0.97	0.92	.86	1.04	.19	.64	1.06	.04	.61
C17	-2.87	.40	-2.68		-2.11	.77	.54	.64	-2.68	.60	.57	-2.65	.05	.56
a	0.34 ^l		0.35		0.34	.86	.46	.55	0.36	.08	.49	0.37	.05	.49
b	0.81 ^l		.84		.81	.96	.35	.36	.84	.18	.54	.85	.04	.53
c	1.07 ^l		.93		1.00	1.27	.82	.81	.96	.13	.44	.98	.04	.42
d	.60 ^l		.63		.63	.98	.50	.55	.64	.13	.52	.66	.03	.50
e	.47 ^l		.53		.55	1.13	.65	.60	.53	.12	.56	.54	.04	.53

TABLE I (continued)

Taft Eq. 1, Parameters ρ		Equation 2, Least Squares Method ^g No constraints (Best C^I, α) ^j (Best C^I, α) ^k			
Reactions ^f	ρ	ρ	α		
Substituents	σ_I^o σ_R^o	ρ_I^M ρ_I^I ρ_I^C ρ_I^M ρ_I^S ρ_I^C ρ_I^M ρ_I^S ρ_I^C	σ_R^o	σ_I^o σ_R^o	
OCH ₃	0.26 -0.42		0.27 -0.43	-0.37	$C^I =$
CH ₃	-.05 -.10		.03 -.18	-.07	0.939
F	.52 -.33		.46 -.29	-.32	$\alpha =$
Cl, Br	.46 -.19		.46 -.18	-.18	0.563
NO ₂	.63 .18		.65 .16	.20	
SD	$\sim \delta_m$	3.95	1.45	3.14	
GOF (DOF) ⁱ		0.182(72)	0.059(30)	0.152(53)	0.138(58)

^a Phenylacetic acids with diphenyldiazomethane, R. M. O'Ferrall and S. I. Miller, J. Am. Chem. Soc., **85**, 2440 (1963).

^b Sapon. of Ethyl trans-2-Phenylcyclopropylcarboxylates, R. Fuchs and J. J. Bloomfield, J. Org. Chem., **28**, 910 (1963).

^c Alk. Hyd. of Ethyl Phenylacetates, R. O. C. Norman and G. K. Redda, J. Chem. Soc., 3247 (1961).

^d Sapon. of Ethyl Phenylpropionates, R. Fuchs, Private Communication to R. W. Taft.

- e Ioniz. of trans-2-Phenylcyclopropylcarboxylic acids, R. Fuchs, C. A. Kaplan, J. J. Bloomfield and L. F. Hatch, J. Org. Chem., 27, 733 (1962).
- f Letter-number code from ref. 1. The reactions so coded are in order of listing; ion-pair formation of benzoic acids with 1,3 diphenyl guanidine; ionization of phenyl acetic acids; ionization of β -phenyl propionic acids; ionization of $\text{ArPO}_2(\text{OH})^-$ in H_2O ; saponification of benzyl acetates; ionization of $\text{ArPO}_2(\text{OH})^-$ in aq. EtOH; condensation of anilines with 2,4-dinitrochlorobenzene. B1 values averaged with those of A. Fischer, B. R. Mann and J. Vaughan, J. Chem. Soc., 1093 (1961).
- g $C^I = \rho_I^P / \rho_I^M$; $C^M = \rho_R^M / \rho_I^M$; $\alpha = \rho_R^M / \rho_R^P$; S is the parameter variance.
- h $C^I = 1$; $C^M = \alpha = 0.5$; Taft's sigmas, cols. 1 and 2 are used.
- i SD = 10^2 (standard deviation); GOF = 10^2 (goodness of fit) = $(10^{-1} \text{SD})^2$. NDAT/DOF; NDAT is the total number of data; DOF is the degrees of freedom, equal to the number of data minus the number of parameters fitted.
- j Taft's σ_I values are used, see col. 1.
- k Unconstrained case σ 's used, see cols. 5 and 6.
- l Hammett-type rhos derived in the original references.
- m SD for the Hammett equation employing the rhos from the original references and the sigmas of D. H. McDaniel and H. C. Brown, J. Org. Chem., 23, 420 (1958).

deviation) and GOF (goodness of fit, defined in footnote i of Table I) measures. Here we obtain the first indication that $\rho_I^M > \rho_I^P$, even though three reactions are seen to exhibit rather strongly inverted values for $C^I (= \rho_I^P / \rho_I^M)$, and interestingly, for two of the three, high values for alpha. That the σ -values agree well with Taft's is to be expected; the latter were used as initial choices for generation of the non-unique, albeit "best", parameter set. [One of an array of parameter sets which fit the experimental data equally well; viz., if all $\rho' = k\rho$ and all $\sigma' = \sigma/k$, the prime set of parameters is identical as regards data fitting to the unprimed set. This example illustrates two of the four degrees of freedom extant in eq. 2 as written, cf., (4).]

3) When equation 2 is constrained to provide the best I and R ratios, the last two sets of unique parameters are obtained. The constraints

are simply that ρ_I^P/ρ_I^M and ρ_R^P/ρ_R^M are constant for all reactions, i.e., equal to C^I and C^R ($=\alpha^{-1}$). The C^I and C^R values providing the best fit to the experimental data are determined in the fitting procedure. The two additional constraints, cf., (4), are imposed by holding σ 's invariant. The SD values correspond to ~ 3 per cent uncertainties and the GOF values are somewhat below that of column 3. These data would seem to suggest that, at least for the σ^o reactivities, the correlational model is as "good" as the data. The last set of results is especially interesting; quite small variances are obtained for the rhos, indicating statistical reliability. The C^M values cluster well about 0.5-0.6. And, perhaps most noteworthy, the best C^I is significantly smaller than unity (a variance on the order of those for the rhos applies here): also, $\alpha > 0.5$.

Including 18 data for NH_2 , OH, CN and non-hydroxylic- NO_2 in the correlation, while holding the previously examined sigmas at their group 2 values, had the following results. No rho was changed by more than 0.03, C^I and α dropped to 0.918 and 0.520 and SD and GOF were, respectively, 3.35 and 0.163 (compare to last group of Table I). The slight decrease in precision of fitting is probably significant and might have been expected (7); 8 of the 18 data added were for non-hydroxylic medium reactions. The σ_I and σ_R values found for NH_2 , OH and CN were: 0.15, -0.57; 0.24, -0.42; 0.57, 0.08. Taft's values are (non-hydroxylic values parenthesized): 0.10, -0.48; 0.16 (0.27), - 0.45 (-0.42); 0.55 (0.48), 0.15.

Further examination of the points raised here is presently underway. The scope of the work is being broadened, as well, to include benzoic acid and other reactivities. Testing of other constraints, e.g., $\rho_R^M = C^M \rho_I^M$ is also being carried out.

It is a pleasure to acknowledge the cooperation of Professor R. W. Taft both for providing compilations of the reactivity data and for several interesting discussions. Several informative discussions were also held

with Dr. W. C. Hamilton of this department concerning statistical methods.

REFERENCES

- (1) Cf., R. W. Taft, Jr. and I. C. Lewis, J. Am. Chem. Soc., 81, 5343 (1959) and other papers in this series.
- (2) Taft claims in ref. 1 to find no significant improvement in the correlations when $I^P \neq I^M$. The relaxation he was able to allow for this constraint was, however, quite limited. Analogy with some cyclo-hexyl reactivities is also cited to support the assumption that $I^P = I^M$. On the other hand, in unpublished work of the author where hypothetical energy cycles are constructed by interchanging substituent and reactant groups, only position dependency of I or strict proportionality of σ_I and σ_R , i.e. $\sigma_I = k\sigma_R$, would allow internal consistency in the Taft equation. See also the related arguments of J. Hine, J. Am. Chem. Soc., 81, 1126 (1959).
- (3) H. O. Hartley, Technometrics, 3, 269 (1959).
- (4) Full details of these methods will be described elsewhere. Both methods in essence involve linearization by initial choices of parameters. In the first, the sigmas are specified whereupon the best rhos for this set choice is obtained. All the rhos are then held constant while the best sigmas are found. The process is continued until no further change in parameters is obtained. In the second method, initial choices are made for all parameters and the differences, D, between observable and computed reaction constants are expanded in a Taylor series. This linearizes the fitting problem; cycling to consistent differences, D, then provides the least-squares optimized parameters. It will suffice here to recognize that unique solutions do not exist for equation 2; at least four constraints must be imposed to allow transformations to unique parameter values. These may be specific in nature, e.g., defined rhos for standard reactions, or general, as is the rho (meta to para) ratio presented above.
- (5) R. W. Taft, Jr., S. Ehrenson, I. C. Lewis and R. E. Glick, J. Am. Chem. Soc., 81, 5352 (1959).
- (6) If $x = 0.05$, $\log [k(1 + x)/k_0(1 - x)] \approx \log (k/k_0) + (2/2.303)(0.05) = \log (k/k_0) + 0.044$.
- (7) Cf., R. W. Taft, E. Price, I. R. Fox, I. C. Lewis, K. K. Andersen and G. T. Davis, J. Am. Chem. Soc., 85, 709 (1963).