Statistical Analysis of Chemical Reactivity Data. Part 5.[1] Aliphatic Substituent Effects

By **Michael Sjöström** * and **Svante Wold**, Research Group for Chemometrics, Department of Chemistry, Umeå University, S–901 87 Umeå, Sweden

We have shown with the aid of principal component analysis combined with a cross-validation technique that a two-term linear free energy relationship (l.f.e.r.) is needed to correlate the data within a set of alkyl-substituted reaction series. It has not been possible to extend the validity of this two-term l.f.e.r. to include CH_2Ph. This substituent in combination with alkyl substituents needs a three-term l.f.e.r. Due to the lack of data it was not possible to confirm if other substituents of the type CH_2X (X = polar group) can be described with the same two-term l.f.e.r. as the alkyl substituents.

SINCE the formulation of the Hammett equation which successfully described the substituent effects for a vast number of reactions of *meta-* and *para*-substituted aromatic compounds, numerous attempts have been made to describe aliphatic reactivity by similar models. From such studies we can discern different viewpoints of the number of substituent parameters that are needed to describe aliphatic reactivity data. For example, Taft [2,3] claims that two effects can be separated, identified as a steric and a polar effect, on the basis of acid- (A) and base- (B) catalysed ester hydrolysis in the following way. A steric substituent scale E_{sX} for a substituent X was defined as the average E_{sX} from three acid-catalysed ester hydrolysis series and two esterification reaction series where equation (1) applies. The rate cofficients

$$E_{sX} = \log(k_X/k_0)_A \qquad (1)$$

k_X and k_O refer to the hydrolysis of XCO_2R' (or esterification of XCO_2H) and CH_3CO_2R' (or esterification of CH_3CO_2H), respectively, with the same R' and reaction conditions. Furthermore, Taft defined the so-called polar scale σ^* using equation (2), *i.e.* as the difference

$$\sigma^* = 0.403[\log(k_X/k_0)_B - \log(k_X/k_0)_A] \qquad (2)$$

between the log relative rate constants for acid- and base-catalysed ester hydrolysis at the same temperature and in the same solvent. For substituents for which complete data were not available, Taft used instead average $\log(k_X/k_0)_A$ and $\log(k_X/k_0)_B$ values mainly from alkaline hydrolysis of ethyl esters in 60 and 70% aqueous acetone and 80% aqueous EtOH.

In contradiction to Taft's analysis, Charton [4-6] has recently claimed that model (3) is sufficient when X is an

$$\log k_X = \psi\nu_X + h \qquad (3)$$

alkyl group. The ν_X constants were interpreted as a steric effect.

Doubts of the significance of the σ^* scale for alkyl substituents have also been expressed by others, reviewed by Shorter,[7] but MacPhee *et al.*[8] have recently claimed that the σ^* scale is valid.

These conflicting results reveal that careful statistical analysis of aliphatic reactivity data is warranted. This can be done by principal component (p.c.) analysis [9,10] combined with the so-called cross-validation (c.v.) technique.[9,11] With p.c. analysis it is possible to determine the parameters α_i, β_{ia}, and θ_{ak} from measurements in a matrix Y (see Figure 1) for model (4) with an

$$y_{ik} = \alpha_i + \sum_{a=1}^{A} \beta_{ia}\theta_{ak} + \varepsilon_{ik} \qquad (4)$$

arbitrary A (a = 1, 2, 3 . . . A). Deviations from the model are expressed by the residuals ε_{ik}. The number of significant terms A in this expression can be determined for a complete data matrix with the c.v. technique. Thus, by means of such data analysis it is possible to determine rigorously whether a one- (*i.e.* Charton's model) or two-component model (*i.e.* Taft's model or perhaps an even more complex p.c. model is needed to describe the data adequately.

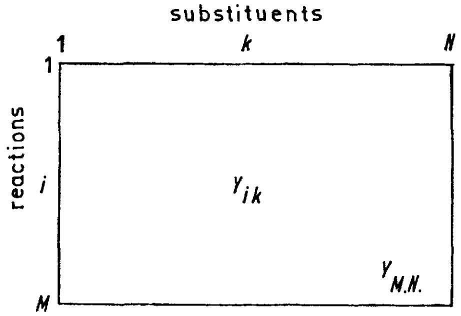

FIGURE 1 Observation matrix Y

We here present a p.c.–c.v. analysis of some data matrices consisting of reactivity data for some alkyl-substituted reaction series.

METHODS

The calculations were done with the statistical package SIMCA [12] designed for p.c. analysis and pattern recognition. The package includes routines for p.c. analysis, c.v., and classification. These methods are well documented in the chemical literature [9,13,14] and here we give just a short description of the modelling-power concept and the classification approach.

To express the variation within a reaction series i, explained by a certain p.c. model, we have used the measure

ψ_i, the so-called modelling-power. This measure is defined by equation (5) where s_i is the residual standard deviation

$$\psi_i = 1 - (s_i/s_{iy}) \tag{5}$$

$$s_i = [M/(M - A) \sum_{k=1}^{N} \varepsilon_{ik}/(N - A - 1)]^{1/2} \tag{6}$$

[equation (6)] for a certain reaction i in a p.c. analysis with A components and where s_{iy} is the data standard deviation [equation (7)] within the reaction series ($A = 0$).

$$s_{iy} = \sum_{k=1}^{N} [(y_{ik} - \bar{y})^2/(N - 1)]^{1/2} \tag{7}$$

Thus a ψ_i value close to unity means that most of the variation within the reaction series is explained by a model with A terms.

Of interest also is to find out how well a certain substituent (index p) not used in the derivation of a certain p.c. model is described by the parameters α_i and β_{ia} from this p.c. model. Such a classification of the data vector y_{ip} for a substituent p can be done by fitting the vector by multiple regression to model (8) with known α_i and β_{ia} values.

$$y_{ip} = \alpha_i + \sum_{a=1}^{A} \theta_{ap}\beta_{ia} + \varepsilon_{ip} \tag{8}$$

θ_{ap} are the regression coefficients. Then the residual standard deviation s_p [equation (9)] is compared with the

$$s_p = \left[\left(\sum_{i=1}^{M} \varepsilon_{ip}{}^2/(M - A)\right)\right]^{1/2} \tag{9}$$

total residual standard deviation for all objects $S_0{}^2$ included in the p.c. analysis by means of the F test [equation (10)]

$$F = s_p{}^2/S_0{}^2 \tag{10}$$

where $S_0{}^2$ is given by equation (11).

$$S_0{}^2 = \sum_{i=1}^{M} \sum_{1=k}^{N} \varepsilon_{ik}{}^2/(N - A - 1)(M - A) \tag{11}$$

We can thus detect whether a substituent p has a significantly higher residual standard deviation or not compared with the typical standard deviation of substituents in a certain model.

RESULTS

A complete data matrix $Y1$ was formed by the reactivity measurements for the reaction series in Table 1 and the following substituents: Me, Et, Pr^n, Pr^i, Bu^n, Bu^i, Bu^t, and cyclo-C_6H_{11}. Two other matrices $Y2$ and $Y3$ were formed by including the complete data vector for CH_2Ph and by excluding the data vector for Bu^t, respectively.

We also considered the data vectors for the substituents Bu^s, cyclo-$CH_2C_6H_{11}$, and $CH(Et)_2$. However, since one measurement each was missing in these vectors (in reactions 3, 5, and 1) we have not used these substituents for the p.c. analysis. For reaction 3 we have used extrapolated rate constants from other temperatures for cyclo-C_6H_{11}, CH_2Ph, and cyclo-$CH_2C_6H_{11}$, and in reaction 4 for cyclo-C_6H_{11}. We are however convinced that these values are as reliable as the other measurements.

P.c.–C.v. Analysis and Classifications.—A cross-validation of $Y1$ showed that to account for the systematic variations within these data a *two*-component model was needed, *i.e.* $A = 2$ in equation (4) (see Table 2). The parameters from this p.c. analysis, α_i, β_{i1}, and β_{i2} are given in Table 3, and θ_{1k} and θ_{2k} for each substituent k are given in Table 4. Table 5 presents the modelling-powers for $A = 1$ and 2 in equation (4) and the standard deviations around the means s_{iy} [equation (7)] for each reaction series i.

A c.v. of $Y2$, the matrix which contained the data vector for benzyl, indicated that a *three*-component model was needed to describe the data (see Table 2). The non-typical behaviour of benzyl compared with the alkyl substituents can also be demonstrated by the classification procedure earlier described [equations (8)—(11)]. The data vector y_i for benzyl was fitted by multiple regression to the parameters α_i, β_{i1}, and β_{i2} from $Y1$. A comparison by means of the F test of the residual variance for benzyl with the typical residual variance for the substituents in $Y1$ revealed significantly larger values for benzyl than ' normal ' (see Table 4).

The data vectors for the substituents with one missing measurement each, namely Bu^s, cyclo-$CH_2C_6H_{11}$, and $CH(Et)_2$ were also fitted to the $Y1$ model with $A = 2$. cyclo-$CH_2C_6H_{11}$ and Bu^s showed typical behaviour but not $CH(Et)_2$ (see Table 4), due to large residuals for reactions 3 and 4.

TABLE 1

Reaction series used in the p.c. analysis

No.	Reaction
1	$RCONH_2 + OH^-$, in H_2O, 75 °C [a]
2	RCO_2H + MeOH, H^+, in MeOH, 50 °C [b]
3	$RCO_2CH_3 + H^+$, in 70% acetone–H_2O, 24.8 °C
4	$RCO_2Et + OH^-$, in 85% EtOH–H_2O, 25 °C [d]
5	pK_a RCO_2H, in H_2O, 25 °C [e]
6	$RCONH_2 + H^+$, in H_2O, 75 °C [f]

[a] P. D. Bolton and G. L. Jackson, *Austral. J. Chem.*, 1969, **22**, 527. [b] H. A. Smith, *J. Amer. Chem. Soc.*, 1939, **61**, 254; 1940, **62**, 1136. [c] G. Davis and P. D. Evans, *J. Chem. Soc.*, 1940, 339; H. A. Smith and J. H. Steele, *J. Amer. Chem. Soc.*, 1941, **63**, 3466; H. A. Smith and R. R. Meyers, *ibid.*, 1942, **64**, 2362. [d] D. P. Evans, J. J. Gordon, and H. B. Watson, *J. Chem. Soc.*, 1937, 1430; H. S. Levenson and H. A. Smith, *J. Amer. Chem. Soc.*, 1940, **62**, 2324; H. A. Smith and H. S. Levenson, *ibid.*, p. 2733. [e] J. F. J. Dippy, *J. Chem. Soc.*, 1938, 1222; M. Kilpatrick and J. D. Morse, *J. Amer. Chem. Soc.*, 1953, **75**, 1846; J. F. J. Dippy and F. R. Williams, *J. Chem. Soc.*, 1934, 161; J. J. Christensen, M. D. Slade, D. E. Smith, R. M. Izatt, and J. Tsang, *J. Amer. Chem. Soc.*, 1970, **92**, 4164; A. J. Ellis, *J. Chem. Soc.*, 1963, 2299. [f] P. D. Bolton and G. L. Jackson, *Austral. J. Chem.*, 1971, **24**, 471.

TABLE 2

R Values from the c.v. of $Y1$ — $Y3$. A component is significant as long as $R < R_{crit.}$

	Components in equation (4)			
Matrix	$A = 1$	$A = 2$	$A = 3$	$A = 4$
$Y1$	0.18	0.08	3.27	0.52
$Y2$	0.25	0.90	0.10	1.34
$Y3$	0.32	0.10	0.93	1.34
$R_{crit.}$	0.94	0.78	0.50	0.22

TABLE 3

Parameters from the p.c. analysis of $Y1$

Reaction i	α_i	β_{i1}	β_{i2}
1	0.749	0.249	−0.525
2	0.809	0.473	−0.114
3	−4.919	0.459	0.052
4	−3.080	0.684	0.460
5	4.864	−0.057	−0.233
6	0.685	0.257	−0.666

TABLE 4

θ_1 and θ_2, F test [equation (10)], and residual standard deviation s_p [equation (9)] for each substituent are given. For substituents denoted (1) the values refer to the p.c. analysis of $Y1$. For a substituent denoted (0) the values refer to when the data vector of the substituent is fitted to the model from $Y1$

Substituent	θ_1	θ_2	F [a]	s_p
Me (1)	1.27	−0.025	1.8	0.046
Et (1)	1.06	−0.197	0.42	0.022
Pr^n (1)	0.454	0.027	0.18	0.015
Pr^i (1)	0.087	−0.111	1.2	0.037
Bu^n (1)	0.427	0.102	3.1	0.061 [b]
Bu^i (1)	−0.786	0.541	0.31	0.019
Bu^t (1)	−2.05	−0.292	0.37	0.021
cyclo-C_6H_{11} (1)	−0.470	−0.047	0.68	0.029
cyclo-$CH_2C_6H_{11}$ (0)	−0.758	0.610	2.5	0.054
CH_2Ph (0)	1.04	0.294	120	0.39 [c]
$CH(Et)_2$ (0)	−2.92	0.940	27	0.18 [d]
Bu^s (0)	−0.942	0.474	1.8	0.047
Bu^t (0) [f]	(−2.22)	(−0.797)	(8.7)	(0.097) [e]

[a] For (1), $F_{crit., 0.05} = 3.3$ and (0), $F_{crit., 0.05} = 2.8$. [b] The largest residual (ε −0.06) in reaction 1. [c] Residuals for reactions 1—6: 0.40, −0.26, −0.37, 0.23, −0.42, 0.02. [d] The largest residuals (0.15 and −0.28) in reactions 3 and 4. [e] The largest residual (0.12) in reaction 1. [f] Classified to the parameters from $Y3$, see text.

TABLE 5

Modelling-powers ψ_i from $Y1$ and s_{iy} [equation (7)] for each reaction series i. The modelling-powers when reaction 5 is excluded are in parentheses

	Reaction					
Model	1	2	3	4	5	6
$A = 1$	0.47	0.91	0.94	0.82	0.15	0.39
$A = 2$	0.84	0.94	0.93	0.97	0.57	0.92
$A = 2$	(0.87)	(0.93)	(0.91)	(0.97)		(0.92)
s_{iy}	0.30	0.51	0.50	0.75	0.09	0.34

For the $Y3$ matrix, as for $Y1$, a two-component model was needed to account for the systematic variation (see Table 2). With $A = 2$, the Bu^t substituent showed a somewhat enlarged residual standard deviation compared with the substituents in $Y3$ (see Table 4). Since Bu^t has an extreme θ_1 value this result is less startling because even very small errors in the model from $Y3$ will result in enlarged residuals. If Bu^t was abnormal the total residual standard deviation S_0 in equation (11) for the model ($A = 2$) from $Y3$, denoted S_{0Y3}, should be much smaller than the corresponding residual standard deviation from $Y1$, S_{0Y1}. This is not the case (S_{0Y1} 0.033 1, S_{0Y3} 0.034 7). This means that the two-component model needed to explain the behaviour of the alkyl substituents in $Y3$ can also explain the behaviour of Bu^t, and no new type of effect is needed to describe this. The enhanced residuals for $CH(Et)_2$ with extreme θ_1 and θ_2 values might (partly) be explained in the same way. This has not been further investigated. The abnormal behaviour of benzyl, described above, cannot be explained in this way since this substituent showed no extreme θ values and since $S_{0Y2} = 0.907$ ($A = 2$), *i.e.* much larger than S_{0Y1}.

Correlation of θ_1 *and* θ_2 *with* E_s *and* σ^*.—For the substituents in $Y1$, except for cyclo-C_6H_{11} for which we have no primary σ^* values, we have investigated the correlation between θ_1 and θ_2 and E_s and σ^*.

Table 6 shows that E_s is highly correlated with θ_1 and that there is no correlation at all between E_s and θ_2. In addition, σ^* is highly correlated with θ_1 as well as with θ_2. We have also investigated E_s and σ^* as functions of θ_1 and θ_2 with the aid of multiple regression (see Table 7). The term $c'\theta_2$ in equation (12) was not significant as expected from the correlation coefficient. Also in equation (13) the term $c''\theta_2$ was not significant at the 95% level (approximately at an 88% level). These results show that E_s and θ_1 contain the same information while σ^* contains basically the same information as a linear combination of θ_1 and θ_2.

TABLE 6

Correlation coefficients

	E_s	σ^*	θ_1
θ_2	−0.015	0.340	−0.012
θ_1	0.999	0.866	
σ^*	0.853		

TABLE 7

E_s and σ^* dependence of θ_1 and θ_2

Model [a]	$F\theta_1$ [b]	$F\theta_2$ [b]	s_e [c]	r [d]
$E_s = a' + b'\theta_1 + c'\theta_2 + e'$ (12)	1 855	0.005	0.03	0.997
$\sigma^* = a'' + b''\theta_1 + c''\theta_2 + e''$ (13)	15	3.9	0.04	0.935

[a] For the multiple regressions in Tables 7 and 8 we have used the SPSS package, N. H. Nie, C. H. Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Brent, 'SPSS: Statistical Package for Social Sciences,' McGraw-Hill, New York, 1975. [b] Analysis of variance for inclusion of θ_1 ($F_{1/5.0.05}$ 6.6) and θ_2 after θ_1 is entered ($F_{1/4.0.05}$ 7.7). [c] Residual standard deviation. [d] Correlation coefficient.

TABLE 8

Reactions 1—6, 2A, and 4A fitted by multiple regression [a] to θ_1 and θ_2 or σ^* and E_S

	$Y_{ik} = \alpha_i + \beta_{i1}\theta_{1k} + \beta_{i2}\theta_{2k} + \varepsilon_{ik}$ (14)				$Y_{ik} = a_i + b_iE_s + c_i\sigma^* + \varepsilon'_{ik}$ (15)			
Reaction	$F\theta_1$ [b,c]	$F\theta_2$ [b,c]	s_i [c,d]	r [e]	F_{Es} [f]	F_σ [f]	s_i [d]	r [e]
1	15	51	0.05	0.994	18.7	1.1	0.16	0.91
2	808	11	0.03	0.9995	775	0.7	0.05	0.997
2A [g]	773	1.3	0.04	0.997	595	0.9	0.05	0.997
3	1 330	0.9	0.04	0.999	4 352	13.8	0.01	0.9999
4	158	167	0.03	0.9998	162	7.9	0.10	0.995
4A [h]	222	0.9	0.12	0.991	196	15.9	0.06	0.998
5	3.5	14	0.04	0.952	0.23	16.4	0.05	0.88
6	10.8	245	0.03	0.998	13.4	3.2	0.17	0.92

[a] See footnote *a*, Table 7. [b] Analysis of variance for inclusion of θ_1 ($F_{1/4.0.05}$ 7.7), and θ_2 after θ_1 is entered ($F_{1/3.0.05}$ 10). [c] The F and s_i values are corrected when θ_1 and θ_2 are calculated from the data. [d] Residual standard deviation, $s_i = [\sum_{k=1}^{N} \varepsilon_{ik}/(N-3)]^{1/2}$. [e] Correlation coefficient. [f] Analysis of variance for inclusion of E_s ($F_{1/5.0.05}$ 6.6) and σ^* after E_s is entered ($F_{1/4.0.05}$ 7.7). [g] Reaction 2 at 25 °C, H. H. Smith and J. Burn, *J. Amer. Chem. Soc.*, 1944, **66**, 1494. [h] Reaction 4 in 70% H_2O–acetone, G. Davis and D. P. Evans, *J. Chem. Soc.*, 1940, 339.

We have also made a comparison when the E_s and σ^* scales and the θ_1 and θ_2 scales are fitted by multiple regression to the reactivity data from reactions 1—6 and two additional reaction series 2A and 4A used by Taft to define E_s and σ^*. These results are presented in Table 8. Except for reactions 3 and 4A the θ_1 and θ_2 scales fit the data better than E_s and σ^*. In the analysis of variance in equation (16) we can see that the θ_1 and θ_2 scales overall fit the reaction series 1—6, 2A, and 4A significantly better than E_s

$$F = \frac{\sum_i s_{i}^{2}{}_{(15)}/(32 - 7)}{\sum_i s_{i}^{2}{}_{(14)}/23} = 3.0,\ F_{\text{crit. }0.05} = 2.0 \qquad (16)$$

and σ^*. $s_{i(15)}$ and $s_{i(14)}$ are the residual standard deviations for the reaction series i in equations (15) and (14) in Table 8. Even in a similar analysis of variance [equation (17)] over the types of reaction series Taft used to define his E_s and σ^* scales (reaction series 2—4, 2A, and 4A) the θ_1 and θ_2 scales performed as well as E_s and σ^*. The reduction of

$$F = \frac{\sum_i s_{i}^{2}{}_{(15)}/(20 - 7)}{\sum_i s_{i}^{2}{}_{(14)}/16} = 1.0,\ F_{\text{crit. }0.05} = 2.4 \qquad (7)$$

the degrees of freedom by 7 in equations (16) and (17) is explained by the fact that at least half the data Taft used to define σ^* and E_s for the alkyl substituents used are those of reactions 2A, 3, 4, and 4A.

DISCUSSION

The c.v. of $Y1$ clearly shows that a two-component model is needed to account for systematic variation within this matrix with alkyl substituents. In a series of papers dealing with the alkyl substituent effects, Charton [4-6] analysed the same reaction series as in this paper, except reaction 5. He claims in contradiction to our findings that equation (3) is sufficient to describe the effect of alkyl substituents in ester and amide hydrolysis and esterification of carboxylic acids. Thus our results reveal that Charton has underestimated the systematic information contained in these data.

We have also found that the calculated θ_1 and θ_2 scales show a resemblance with Taft's E_s and σ^* scales; however our scales where the individual properties of the measurements are minimized, overall fit the investigated data better than E_s and σ^*.

For the reaction series 1—6 few measurements are available for substituents of the type CH_2Y where Y is a polar group (Y = halogen, OR, CN, NO_2, *etc.*). This meant that we could not elucidate whether CH_2Y substituents in general are on the same scales as alkyl substituents or if they behave like benzyl which could not be described by the same model as alkyl substituents. Our results indicate strongly, however, that a *general* linear free energy relationship for aliphatic reactivity must include at least three terms.

From the modelling-powers for $Y1$ (see Table 5) we can see that the model with $A = 2$ describes most of the variation within the reaction series 1—6. The only exception is reaction 5, ionization of carboxylic acids, for which only slightly more than half the variation within the reaction series is explained by the model. We have also found that if Bu^t is excluded ($Y3$) the modelling-power for this series drops to 0.20 ($A = 2$) which means that the model described almost none of the variation in pK_a for the lower alkyl substituted carboxylic acids. Since the standard deviation s_Y [equation (7)] is much smaller for reaction 5 compared with the other reaction series, the inclusion of this series affects the fit very little as can be seen for a comparison of the modelling-powers with and without reaction 5 (see Table 5). The conclusions of normal behaviour for Bu^t and abnormal behaviour for benzyl also hold if reaction 5 is excluded.

A separate problem from the determination of the number of components in equation (4) and the calculations of the parameters in this model, is the interpretation of the derived substituent scales as measures of steric, polar, *etc.*, effects. Such interpretations must be based on some external assumptions, like those on which Taft based his interpretation of E_s and σ^*. If his assumptions that the variation in acid- and base-catalysed ester hydrolysis is due to steric and polar effects then also θ_1 can be considered a measure of a steric effect, and a linear combination of θ_1 and θ_2 a measure of a polar effect.

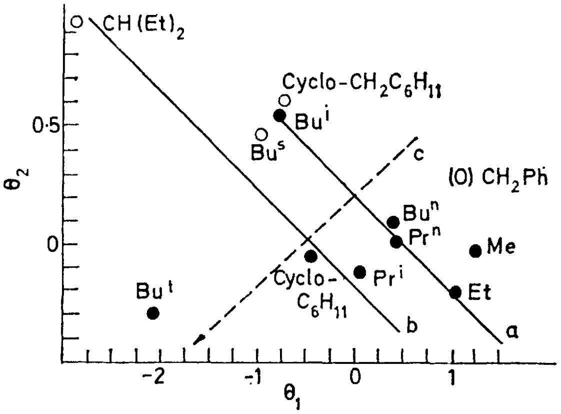

FIGURE 2 θ_1 plotted against θ_2. Full circles refer to the θ values from the p.c. analysis of $Y1$ and open circles from the classification of the data vectors for substituents with missing data to the model from $Y1$. Substituents with one and two α-hydrogens are extended along two separate lines *a* and *b*. The degree of substitution at the α-carbon increases along axis *c*

However, it has been shown [15] by Taylor expansions of a function $F = y_{ik}$ that a model with $A = 2$ cannot give a unique interpretation of the number of substituent effects that varies within F. With such Taylor expansions it is possible to show that models with $A = 2$ can also be obtained even with one type of substituent effect that varies within the data set, but the reaction dependent effect within the series varies too much (caused for example by mechanistic complications). If we approve this possibility, then we must also accept that from a chemical point of view the interpretation of these results is more difficult. Just to assume two effects and even interpret the nature of these effects seems in our view to be questionable.

It is of interest to study a plot of θ_1 against θ_2 (see Figure 2). This plot shows that the substituents with one and two α-hydrogens are extended along two separate but approximately parallel axes denoted *a* and *b*. Perpendicular to these axes we can see that the degree of substitution increases (*c*). This indicates that the alkyl substituents with one and two α-hydrogens can be well described with two separate models with $A = 1$.

In the p.c. analysis we have only studied reactions with substituents at the acyl site. Like earlier investigators [2,16] we have found that scales from the acyl part cannot describe well the behaviour of the alkyl part in, for example, alkaline hydrolysis of alkyl acetates. When such a series is fitted to the θ_1 and θ_2 scales, the modelling-power is 0.75 and Bu^i and Bu^s especially deviate strongly. This together with the low modelling-power for the carboxylic acids indicates the inadequate reference frame for the present θ_1 and θ_2 scales. This limitation also exists for all other one- or two-parameter substituent scales for aliphatic reactivity and will exist as long as no investigation is made with the aim of producing a more complete reactivity data matrix for a large number of substituents and relevant aliphatic reactions.

The project is supported by grants from the Swedish Natural Science Research Council.

[8/2053 *Received, 27th November*, 1978]

REFERENCES

[1] Part 4, M. Sjöström and S. Wold, *Chemica Scripta*, 1976, **9**, 200.

[2] R. W. Taft, *J. Amer. Chem. Soc.*, 1952, **74**, 2729.

[3] R. W. Taft in 'Steric Effects in Organic Chemistry,' ed. M. S. Newman, Wiley, New York, 1956.

[4] M. Charton, *J. Amer. Chem. Soc.*, 1975, **97**, 1552.

[5] M. Charton, *J. Amer. Chem. Soc.*, 1975, **97**, 3691.

[6] M. Charton, *J. Org. Chem.*, 1976, **41**, 2906.

[7] J. Shorter in 'Advances in Linear Free Energy Relationships,' eds. N. B. Chapman and J. Shorter, Plenum Press, London, 1972.

[8] A.-J. MacPhee and J.-E. Dubois, *Tetrahedron Letters*, 1976, 2471.

[9] S. Wold and M. Sjöström in 'Chemometrics, Theory and Practice,' ACS Symposium Series No. 52, ed. B. R. Kowalski, Amer. Chem. Soc., Washington, 1977.

[10] S. Wold, *Pattern Recognition*, 1976, **8**, 127.

[11] S. Wold, *Technometrics*, 1978, **20**, 397.

[12] S. Wold, SIMCA-2T manual, available from the author.

[13] S. Wold and M. Sjöström in 'Correlation Analysis in Chemistry: Recent Advances,' eds. J. Shorter and N. B. Chapman, Plenum Press, London, 1978.

[14] M. Sjöström and U. Edlund, *J. Magnetic Resonance*, 1977, **25**, 285.

[15] S. Wold, *Chemical Scripta*, 1973, **6**, 97.

[16] M. Charton, *J. Org. Chem.*, 1977, **42**, 3531.