CHEMOMETRIC ANALYSIS OF SUBSTITUENT EFFECTS.
I. SUBSTITUENT EFFECTS ON THE DISSOCIATION
OF MONOSUBSTITUTED BENZOIC ACIDS

Oldrich PYTELA

Department of Organic Chemistry,
University of Chemical Technology, 532 10 Pardubice, The Czech Republic

Forty-six representative sets of data relating to the dissociation of monosubstituted benzoic acids in various solvents were extracted from the literature. The set of substituents included 25 common substituents in the *meta* position and the same number of substituents in the *para* position. Hydrogen served as the reference standard. The sets were subjected to regression analysis using conventional empirical models. The Hammett model was found to be valid within the limits of experimental error. The Taft model with the σ_I, σ_R^0, σ_R^+ and σ_R^- parameters is the best model to account for the substituent effects from the *meta* or *para* position solely. The inductive and mesomeric effects of the substituent are also best separated on this parametric scale. By applying the method of conjugated deviations (analysis of latent variables), a single latent variable was found to be sufficient to describe the data variability in all the three data sets analyzed (*meta + para* in the Hammett model and sets of *meta* and *para* substituted derivatives separately). The relationship between the first latent variables from the *meta* and *para* positions is isoparametric, the substituents lie on three straight lines intersecting in one point. The first straight line corresponds to substituents with the *I* effect (CH$_3$, C$_2$H$_5$, *tert*-C$_4$H$_9$, C$_6$H$_5$, SO$_2$NH$_2$, CN, NO$_2$ and hydrogen as the reference standard), the second straight line corresponds to substituents with *I* and +*M* effects (NH$_2$, N(CH$_3$)$_2$, NHCOCH$_3$, CH$_3$O, SH, F, Cl, Br), and the third straight line corresponds to substituents with *I* and −*M* effects (CHO, CH$_3$CO, COOR, SO$_2$CH$_3$, CF$_3$). The +*M* mesomeric effect is twice as strong as the −*M* effect. These facts were used to propose a new empirical model for the description of substituent effects by means of one substituent constant and one (*meta*) or two (*para*) reaction constants. The PLS method revealed that the additional effects contribute about 8% to the data variability in the interpretation of the *para* substitution through *meta* substitution.

Similarity of response to an introduced perturbation of the same kind is one of the most significant properties of chemical systems. The underlying physico-chemical principle is based on the similarity of changes in the Gibbs energies[1–11] (LFER, ETR principle). Mathematical description consists in a linear combination of general parameters, which are adjusted on a chemical model or emerge from statistical processing of large sets of experimental data. As is the case with other phenomena such as the solvent effect and nucleophilicity, substituent effects, also, are basically examined by verifying the extent of validity of the linear model and determining the necessary number of explanatory

parameters[7]. The above-mentioned similarity of system response – which is a similarity of substituent effects on the reaction center in the particular case of substituent effects – thus constitutes the basic condition for linearity. If this is satisfied, one can always find a set of parameters describing the nature of the perturbation (the substituent) for which the resulting effect–substituent relationship is described by a linear form or at least a nearly linear form in relation to the extent of similarity[7,10,11]. The linear dependence fit is a function of the number of explanatory parameters. Methods of analysis of latent variables[12] (PCA, FA, PLS, etc.) are currently best suited to the determination of the number of parameters in linear models. Using published sets of substituent constants in empirical equations of several types, authors have arrived at numbers of parameters equal to two[13 – 15], three[16,17], four[18], five[19] and other[20]. As expected, calculations based on homogeneous sets of empirical data lead to lower numbers of latent variables, viz. one[21,22], two[22 – 27] or three[26 – 29]. The differences in the substituent effects from the *meta* and *para* positions are also marked, the substituent effect transfer being more complex from the *para* position than from the *meta* position[30]. The variety of the results cited is a consequence of the different kinds, extents, homogeneity and quality of data sets subjected to the analysis. Apart from extreme cases, the number of significant parameters seems to be no higher than three. The physico-chemical meaning of the established latent variables is usually elucidated via substituent constants derived from experimental models[16,21 – 24,29]. The results of the above studies which are based on isolation of the latent variables, as well as those of studies by regression and correlation analysis of substituent constants[31 – 47] and by other approaches[41,48 – 53] demonstrate that there exist two substituent effect transfer pathways, which are more or less independent: spatial transfer via π bond electrons, and spatial transfer via σ bond electrons. Orthogonality or nonorthogonality of the pathways is determined by the extent of overlap (geometry of the molecule) and energy difference of the corresponding molecular orbitals. Specific and nonspecific solvation[54 – 58] contribute as supplementary effects; steric interactions[59,60] are a specific phenomenon. The extent of the substituent effect transfer to the reaction center is apparently also affected by the mutual interactions between the substituent and the reaction center, occurring via transfer channels[43]. The above solvent interactions are usually described by means of additional substituent parameters[58] or by modifications of the existing parameters[57], whereby the description of the substituent effects is additionally complicated. In comparison with parameter modification, addition of another term appears preferable because such a more complex model can always be reduced to the basic model. Equations with too many explanatory variables, however, must be rejected both in theory and practice because the intercorrelations (which are quite extensive in some cases) mostly lead do misinterpretation for low numbers of independent variables (substituents). Substituent classification[14,15,17,33,61] deserves particular attention. The most extensive study of this kind,

based on 29 rather heterogeneous descriptors, resulted in a classification of substituents into four classes, viz. alkyls, halogens, donors and acceptors[61].

Two kinds of correlation equations are used to describe substituent effects. The one approach, historically older, is based on the assumption of an identical extent of transfer of the effects between the reaction center and the substituent and different nature of substituents in the *meta* and *para* positions. This type is represented by the Hammett equation[1] in an appropriate parametrization[7,10,11],

$$\Delta G = \Delta G^0 + \rho\,\sigma_{m,p}. \tag{1}$$

The insufficient universality of Eq. (*1*) led to the proposal of a different kind of correlation equations, respecting the separation of the description of the various effects of the substituent (σ) and extent of interaction of such effects with the reaction center (ρ), irrespective of the site of substitution. Generally, such correlation equations can be written as

$$\Delta G = \Delta G^0 + \rho_1\,\sigma_1 + \rho_2\,\sigma_2 + \rho_3\,\sigma_3 + \ldots, \tag{2}$$

the number of terms being usually two or three[30,62]. Correlation equations involving the inductive and resonance substituent effects[30,63] or other physical[40] and statistical[43] characteristics are typical representatives. The additive-multiplicative model with a cross-term as a third term, which is conventionally used to account for the interaction of two substituents[64 – 68], is a variant of regression equations with three explanatory parameters. This model acquires an entirely different interpretation if it is construed as a description of the effect of a single substituent with two effects, described by two parameters, σ_1 and σ_2. Provided that this model is valid, the mixed term mirrors the extent of orthogonality of the parametrically described effects.

In view of the complexity of the substituent effects, chemical models in which side effects are suppressed as much as possible are of fundamental importance for the study and description of the effects. Rigid compounds possessing a well-defined geometry and properties are particularly well suited. Benzoic acid and its functional derivatives hold a prominent position among such substances. Many publications have been devoted to the dissociation of *meta* and *para* substituted benzoic acid derivatives (vide infra and refs[58,69 – 71]), less attention has been paid to IR spectroscopic investigation (e.g. refs[72,73]) and NMR investigation (e.g. benzonitriles[22]). The dissociation of benzoic acid has also been treated by quantum chemical methods[74 – 77] and the theoretical dissociation constants match closely the observed gas phase data[78].

Although so extensive experimental material concerning benzoic acid exists, it has never been applied in its entirety to the study of substituent effects. The present paper

is therefore devoted to a chemometric analysis of sets of dissociation constants of *meta* and *para* substituted benzoic acids in various solvents. This includes interpretation of the results by various parametric models and analysis of the latent variables, determination of their number and elucidation of their physical meaning in the various empirical models.

THEORETICAL

Twenty-five substituents in the *meta* position, the same number of substituents in the *para* position, and hydrogen as a substituent were chosen (Table I). Dissociating substituents ($COOH$, SO_3H) and substituents entering into appreciable interaction with the solvent (OH) were excluded. The basic set of data extracted from the literature comprised logarithms of dissociation constants or related quantities for dissolved *meta* and *para* substituted benzoic acids (Table II). The choice was made so that the total number of substituents was never lower than 9. This basic set was divided into a set of *meta* substituted derivatives containing at least 6 substituents (all of the sets in Table II except No. 3 and 46) and a set of *para* substituted derivatives containing no less than 9 members (sets No. 1, 5, 6, 9, 10, 12 – 15, 17, 19, 23, 24, 26, 41 – 46 in Table II). The data were arranged into matrices whose rows corresponded to substituents in the order in Table I and columns corresponded to the individual experiments (solvents in

TABLE I
Substituents chosen for analysis

No.	Substituent	No.	Substituent	No.	Substituent	No.	Substituent
1	H						
2	3-CH$_3$	15	3-N=NC$_6$H$_5$	27	4-CH$_3$	40	4-N=NC$_6$H$_5$
3	3-C$_2$H$_5$	16	3-NO$_2$	28	4-C$_2$H$_5$	41	4-NO$_2$
4	3-*tert*-C$_4$H$_9$	17	3-OCH$_3$	29	4-*tert*-C$_4$H$_9$	42	4-OCH$_3$
5	3-C$_6$H$_5$	18	3-OCOCH$_3$	30	4-C$_6$H$_5$	43	4-OCOCH$_3$
6	3-CF$_3$	19	3-SH	31	4-CF$_3$	44	4-SH
7	3-CCl$_3$	20	3-SCH$_3$	32	4-CCl$_3$	45	4-SCH$_3$
8	3-CN	21	3-SO$_2$CH$_3$	33	4-CN	46	4-SO$_2$CH$_3$
9	3-CHO	22	3-SO$_2$NH$_2$	34	4-CHO	47	4-SO$_2$NH$_2$
10	3-COCH$_3$	23	3-F	35	4-COCH$_3$	48	4-F
11	3-COOR	24	3-Cl	36	4-COOR	49	4-Cl
12	3-NH$_2$	25	3-Br	37	4-NH$_2$	50	4-Br
13	3-N(CH$_3$)$_2$	26	3-I	38	4-N(CH$_3$)$_2$	51	4-I
14	3-NHCOCH$_3$			39	4-NHCOCH$_3$		

Table II). The substituents 3-CCl$_3$, 3-N=NC$_6$H$_5$, 3-OCOCH$_3$, 3-SCH$_3$, 4-CCl$_3$, 4-N=NC$_6$H$_5$ and 4-OCOCH$_3$ were eliminated from the sets processed.

The sets were interpreted as a whole using the conventional empirical correlation equations with the substituent constants σ_m, σ_p, σ_p^0, σ_p^+, σ_p^-, σ_I, σ_R^0, σ_R^+, σ_R^- (ref.[7]), σ_χ, σ_α, σ_F, σ_R (ref.[40]) and σ_l, σ_d, σ_e (ref.[43]). Summary statistical characteristics of the experimental set as a whole, taking into account the statistical significance of the regression coefficients in the individual regressions, were employed to assess the interpretation significance of the various correlation relationships. The residual standard

TABLE II
Selected sets of experimental data on the dissociation of *meta* and *para* substituted benzoic acids (at 298 K unless stated otherwise); n is the number of derivatives in the set

No.	Solvent	n	Ref.	No.	Solvent	n	Ref.
1	water	30	69	26	80% methyl cellosolve	21	86
2	water	13	79	27	80% methyl cellosolve	13	88
3	water[a]	10	80	28	20% dioxane[b]	11	89
4	water[a]	14	81	29	40% dioxane[b]	11	89
5	water, tetrabutylammonium bromide	17	82	30	41.5% dioxane	15	90
6	methanol	31	69	31	55% dioxane	15	90
7	methanol	11	83	32	60% dioxane[b]	11	89
8	10% ethanol	12	84	33	71.5% dioxane	15	90
9	10% ethanol	25	85	34	80% dioxane[b]	11	89
10	10% ethanol	17	58	35	83% dioxane	15	90
11	50% ethanol	13	84	36	dioxane–water $\varepsilon = 55$	11	83
12	50% ethanol	35	85				
13	50% ethanol	21	86	37	dioxane–water $\varepsilon = 40$	11	83
14	50% ethanol	18	87				
15	50% ethanol	34	58	38	dioxane–water $\varepsilon = 15$	11	83
16	75% ethanol	13	84				
17	75% ethanol	34	58	39	10% acetone	13	79
18	85% ethanol	21	58	40	25% acetone	13	79
19	ethanol	31	69	41	acetone	29	69
20	ethanol	11	83	42	acetonitrile	28	69
21	1-propanol	11	83	43	dimethylformamide	31	69
22	1-butanol	11	83	44	tetramethylene sulfone	28	69
23	22% 2-methyl-2-propanol	27	58	45	dimethyl sulfoxide	19	91
24	32% 2-methyl-2-propanol	34	58	46	nitromethane	13	92
25	1,2-ethanediol	11	83				

[a] Data in ΔG, kcal mol^{-1}; [b] at 303 K.

deviation of the whole set was derived from the differences between the observed and calculated values of the dependent variable e_{ij} in the individual regressions according to the formula

$$s = \left[\frac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} e_{ij}^2}{\sum\limits_{j=1}^{m} n_j - \sum\limits_{j=1}^{m} P_j} \right]^{1/2} , \qquad (3)$$

where n and m are the numbers of rows and columns, respectively, of the matrix, n_j is the number of values in the j-th column (given in Table II for the basic set) and P_j is the number of statistically significant regression parameters in the individual regressions. The analysis of the latent variables in the sets relied on the conjugate deviations method (CDA, ref.[93], starting decomposition was made following ref.[65]) with parameters correction according to the modified relation

$$\Delta P_k = \frac{\sum\limits_{i=1}^{n} (y_i - \hat{y}_i) \, |a_k|}{a_k \sum\limits_{p=1}^{P} |a_p|} . \qquad (4)$$

The relation between the sets with substituents in the *meta* and *para* positions was analyzed by the PLS method. The residual standard deviation was invariably determined according to the formula[93]

$$s = \left[\frac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} e_{ij}^2}{\sum\limits_{j=1}^{m} n_j - n \, P - \sum\limits_{j=1}^{m} P_j} \right]^{1/2} , \qquad (5)$$

where the symbols are as in Eq. (3); P is the number of latent variables. The explained variability fraction V, multiple correlation coefficient R and number of significant latent variables were determined as described in ref.[93]. All calculations used standardized data to enable results from different experiments to be compared.

RESULTS AND DISCUSSION

Regression Analysis

Application of Eqs (*1*) and (*2*), with conventional parametric substituent constant sets, to the data of Table II gave summary statistical characteristics which are given in Table III. The results demonstrate that there is no statistically significant difference between the residual standard deviations in the various best-parametrization empirical models. As expected, benzoic acid as a standard chemical model satisfies the Hammett equation (*1*), and the use of more general models (*2*) contributes no interpretation of additional data variability. The correlation fit in the interpretation of the substituent effects from the *meta* position is apparently best using sets of parameters derived from the Hammett equation[7] and its modifications, or from the DSP dual principle[63]. The substitution constants σ_p (or σ_p^0, σ_p^+ and σ_p^-), if deliberately employed as regressors,

Table III
Summary standardized residual standard deviations s (Eq. (*3*)), total explained variability V, total multiple correlation coefficients R and degrees of freedom ν in the empirical models (*1*) and (*2*) for the parametric substituent constant sets used

Model and constants	s	V, %	R	ν
Equation (*1*)		*meta + para*		
$\sigma_m + \sigma_p$	0.188	96.9	0.984	717
Equation (*2*)		*meta*		
σ_m, σ_p	0.209	96.7	0.983	325
$\sigma_m, \sigma_p^0, \sigma_p^+, \sigma_p^-$	0.212	97.4	0.987	250
$\sigma_I, \sigma_R^0, \sigma_R^+, \sigma_R^-,$	0.216	97.5	0.987	229
$\sigma_\chi, \sigma_\alpha, \sigma_F, \sigma_R$	0.715	72.8	0.853	185
$\sigma_l, \sigma_d, \sigma_e$	0.457	87.0	0.933	258
Equation (*2*)		*para*		
σ_m, σ_p	0.214	96.8	0.984	177
$\sigma_m, \sigma_p^0, \sigma_p^+, \sigma_p^-$	0.229	97.9	0.989	101
$\sigma_I, \sigma_R^0, \sigma_R^+, \sigma_R^-,$	0.250	97.6	0.988	95
$\sigma_\chi, \sigma_\alpha, \sigma_F, \sigma_R$	0.526	88.5	0.941	105
$\sigma_l, \sigma_d, \sigma_e$	0.414	91.4	0.956	126

were statistically significant in a high number of experiments analyzed, occasionally their combination even replaced the substitution constant σ_m. This is indicative of a linear relation between σ_m and σ_p, as demonstrated recently[30]. On the other hand, the substituent effects are well separated on the σ_I, σ_R scale. In the vast majority of cases, the substituent effect in the *meta* position was interpreted, in addition to the σ_I constant, also by some of the constants σ_R^0, σ_R^+, σ_R^-. The correlation fit in the interpretation of substituent effects from the *para* position is identical with that for the *meta* position. Using the σ_m, σ_p set the substituent constant σ_m plays a role as a regressor in more 3 cases out of 20, always as a supplementary, less significant component. If the extended set of σ_m, σ_p^0, σ_p^+, σ_p^- is used, σ_p^+ becomes the most significant regressor, the remaining substituent constants being involved with a roughly identical frequency. Interpretation by means of the σ_I, σ_R^0, σ_R^+, σ_R^- set was invariably expressed as a combination of the inductive and resonance effects. The latter was most frequently expressed as σ_R^+ and least frequently as σ_R^-.

From the above facts it follows that the substituent effects affecting the dissociation of benzoic acid are involved to roughly the same extent on the σ_m, σ_p (σ_p^0, σ_p^+, σ_p^-) scale as well as on the σ_I, σ_R (σ_R^0, σ_R^+, σ_R^-) scale; probably, they can be mutually transformed into one another. Separation of the individual effects is unambiguously better on the general scale of σ_I, σ_R, which thus emerges as more suitable from the theoretical as well as interpretation point of view. The extent of unexplained variability is about 3%, irrespective of medium used. The analogous calculations for the dissociation constants measured in water (experiments No. 1 – 4 in Table II) provide a value of unexplained variability about 0.3%. This difference is partly due to the lower accuracy of dissociation constant measurement in nonaqueous solvents; for the most part, however, to the use of substituent constants adjusted in water as the standard solvent. In this manner, changes in the specific solvation of the substituent, in the specific solvation of the reaction center associated with changes in its back-effect on the substituent, as well as changes in the non-specific effects (dielectric properties, polarizability) on the entire interacting system are disregarded. For benzoic acid as a chemical model of substituent effects, all of these effects are smaller than or comparable to the experimental error.

Analysis by the Method of Conjugated Deviations (CDA)

The results of application of this method are summarized in Table IV. Comparison of the standard deviation of a latent variable (t_{mp}) with that of the Hammett data model in Table III demonstrates a very good quality of the parametric set used[7], although the difference is no more statistically insignificant ($F(717,717) = 1.25$, $F_{0.975} = 1.16$). Unexplained variability is roughly comparable to the experimental error. This also applies to the sets of *meta* and *para* substituted derivatives. It can be thus concluded that the model with one latent variable is valid for the dissociation of benzoic acid within the experimental precision regardless of solvent, so that there is no point in adding another

latent variable. This is consistent with published data[30]. The correlation of the first latent variable with the Hammett substituent constants is fairly tight ($r = 0.995$). The t_{mp} values for 3-SH, 4-NHCOCH$_3$, 4-SH and 4-F substituents were outliers on the Jack-knife residuals, and the correlation improved ($r = 0.997$) when these were eliminated; the t_{mp} value for the CHO group then emerged as a next outlier. On the whole, the relation between the first latent variable, t_{mp}, and the Hammett substituent constant scale[7] is tight, which indicates a good parametrization. The first latent variable in the set of *meta* substituted derivatives (sets No. 26, 29 and 37 in Table II were omitted due to their low modelling power) gives the closest correlation with the Hammett constant σ_m (ref.[7], $r = 0.996$), as in the correlations with the individual sets in Table III. The Jack-knife residuals indicated the t_m values for the 3-CHO and 3-SH substituents as outliers. After eliminating them, the correlation improved slightly ($r = 0.998$) and 3-*tert*-C$_4$H$_9$ gave a next t_m outlier (after its elimination, $r = 0.998$). The correlation with the substituent constants σ_I and σ_R^0 (ref.[40]) was looser ($R = 0.987$) but it improved to a value comparable to that for σ_m if the outliers for 3-CHO and 3-SH ($R = 0.987$) and additionally for 3-F ($R = 0.998$) were left out. The correlations with the sets of σ_F, σ_R (ref.[40], $R = 0.984$; without 3-SH, $R = 0.989$) and σ_l, σ_d, σ_e (ref.[43], σ_l, σ_d solely, $R = 0.982$; without 3-SH, $R = 0.995$) were least successful. The indication of outliers can be due to the t_m value, which seems likely for the substituent 3-SH according to the CDA standard deviations, or to incorrect substituent constant values, which seems to be the case with the 3-CHO substituent (underestimated). As in the correlations of the initial data with the various substituent constant sets (Table III), the statistical significance of the terms accounting for the resonance contribution is surprising. There are only two feasible explanations: either this effect is also transferred from the *meta* positions, or the inductive, mesomeric and other effects, if any, are incorrectly separated in the substituent sets; the latter explanation seems more likely.

TABLE IV
Summary standardized residual standard deviations s (Eq. (5)), explained variability V and degrees of freedom ν in conjugated deviations analysis (CDA) applied to the data in Table II for the various numbers of latent variables p

p	*para + meta*			*meta*			*para*		
	s	V, %	ν	s	V, %	ν	s	V, %	ν
1	0.161	97.72	717	0.173	97.48	342	0.162	87.86	206
2	0.131	98.66	640	0.142	98.50	303	0.119	99.04	170
3	0.116	99.07	568	0.120	99.07	265	0.098	99.50	131

The first latent variable in the set of *para* substituted derivatives gives the tightest correlation with the Hammett substituent constant σ_p (ref.[7], $r = 0.994$), as in the correlation with the individual data sets (Table III). None point was identified as an outlier. The correlation for the σ_F set was somewhat poorer (ref.[40], $R = 0.991$) and the t_p values for 4-NH$_2$ and 4-SH were outliers. The correlation improved on eliminating them ($R = 0.996$) and no other value was found to be an outlier. The use of the σ_l, σ_d, σ_e set (ref.[43]) gave poorer correlations ($R = 0.984$, or $R = 0.992$ on eliminating 4-SH), and the situation with the σ_I and σ_R^0 set was similar (ref.[40], $R = 0.984$, or $R = 0.986$ on eliminating 4-NH$_2$). The frequent identification of the SH group as an outlier, irrespective of its substitution site, corresponds with the high standard deviation in the method of conjugated deviations. Presumably, this is due to the specific solvation of this group, as is the case with the OH group. The contributions of the terms accounting for the inductive and resonance effects are roughly comparable, with a better link to the description of the resonance contribution, as documented by the partial correlation coefficients

$$r_{t\sigma_F \cdot \sigma_R} = 0.964, \qquad r_{t\sigma_R \cdot \sigma_F} = 0.978$$

$$r_{t\sigma_l \cdot \sigma_d \sigma_e} = 0.948, \qquad r_{t\sigma_d \cdot \sigma_l \sigma_e} = 0.964, \qquad r_{t\sigma_e \cdot \sigma_l \sigma_d} = 0.486$$

$$r_{t\sigma_I \cdot \sigma_R} = 0.943, \qquad r_{t\sigma_R \cdot \sigma_I} = 0.969.$$

Of interest is the relation between the first latent variables obtained from the sets of *meta* (t_m) and *para* (t_p) substituted benzoic acid derivatives, as shown in Fig. 1. This plot demonstrates that the set of substituents decomposes into three lines (a bundle of
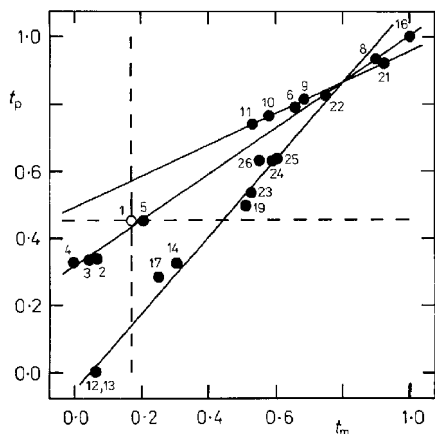

Fig. 1

Distribution of substituents in the plane of the first latent variables t_m and t_p

straight lines) which intersect in a common point. This bundle can be described by Eq. (*6*)

$$t_p = t_p^0 + (I + \delta\Delta M)(t_m - t_m^0),\qquad(6)$$

where t_p^0 is the t_p value for $t_m = t_m^0$, I is a measure of sensitivity to the inductive effect, ΔM is a correction for the mesomeric effect, and δ is a parameter which adopts values of $\delta = 0$ for substituents with an inductive effect solely, $\delta = +1$ for substituents with an inductive and a positive mesomeric effect, and $\delta = -0.5$ for substituents with an inductive and a negative mesomeric effect. The values of $\delta = 0$ and $\delta = +1$ were postulated, the value of $\delta = -0.5$ emerged as a rounded estimate from optimization. Each substituent for which the t_m and t_p values were calculated was assigned one of the δ values (classification), and the decomposition into the classes was optimized by minimization of the residual standard deviation in Eq. (*6*). Substituent No. 26 (iodine) was subsequently eliminated from the calculations as a transition substituent.

In the resulting classification, Class 1 of substituents with the I effect ($\delta = 0$) comprises the substituents CH_3, C_2H_5, *tert*-C_4H_9, C_6H_5, SO_2NH_2, CN and NO_2; H as the reference standard is also included. The inclusion of the cyano and nitro groups is a surprise from the point of view of the conventional Hammett substituent constants, adjusted based on measurements in water. However, measurements of the dissociation constants of benzoic acid in organic acids revealed[69] that the effects of the nitro and cyano groups from the *meta* and *para* positions are not very different[57]. The solvent effect is eliminated from the latent variables analyzed, and in view of the results obtained, the different effect of the NO_2 and CN groups from the *meta* and *para* positions as observed in water is due either to the nonspecific solvent effect (relative permittivity in relation to the dipole of the molecule) or to the specific solvation[58] (solvent acidity). Class 2, which contains substituents with I and $+M$ effects ($\delta = +1$), comprises NH_2, $N(CH_3)_2$, $NHCOCH_3$, CH_3O, SH, F, Cl and Br, as expected. Exhibiting anomalous behaviour, iodine as a substituent lies somewhere between Classes 1 and 2. The methoxy and acetylamino groups also provide a somewhat lower contribution (Fig. 1). Class 3, which contains the lowest number of members, comprises substituents in which the X=Y bond is polarized appreciably due to highly different electronegativities of the two atoms (difference in electronegativities exceeding one). The substituents CHO, CH_3CO, COOR, SO_2CH_3 and, due to hyperconjugation, apparently also CF_3 are included. A notion of the nature of these substituents can be gained from Fig. 1, where the substituents are situated in quadrants with hydrogen in the center. Quadrant I contains electron acceptors with the $-I$ effect (Class 1), with $-I$, $+M$ effects ($-I$ prevailing, Class 2), and with $+I$, $-M$ effects (Class 3). Quadrant II involves no substituent, that is, no substituent with $-I$ and $-M$ effects exists. Substituents in quadrant III are electron donors with $+I$ effect (Class 1) and with $+I$, $+M$ effects (Class 2). No substituent with

+I and −M effects is included. The last quadrant, IV, includes electron donor substituents with −I, +M effects solely (+M prevails, Class 2). The decomposition of substituents into three straight lines (classes) with a common point of intersection evokes another interesting idea, namely, that in principle there may exist substituents with −I and +M effects that will be stronger electron acceptors than substituents with, e.g., −I and −M effects. Similar is the case of substituents with −I effect (CN, NO$_2$), which are stronger electron acceptors than substituents with −I and −M effects (SO$_2$CH$_3$). The classification suggested by us is not entirely consistent with published classifications[14,15,17,33,61] but it approaches closely the division of substituents in the plane of the substituent constants σ_I and σ_R (ref.[33]). The differences found seem to be due the fact that a single chemical model has been analyzed in this work whereas rather heterogeneous data have been subject to analysis as reported in the literature cited.

Using the substituent classification, the validity of model (6) was examined by testing the ratio of the residual variances in this model and in the model with three independent straight lines. The following residual sums of squares were obtained: $S_R = 8.314 \cdot 10^{-2}$ for the straight line bundle and $S_R = 2.057 \cdot 10^{-2}$, $6.148 \cdot 10^{-2}$ and $6.471 \cdot 10^{-2}$ for Classes 1, 2 and 3, respectively. The residual variance was $s^2 = 4.89 \cdot 10^{-3}$ for the bundle and $5.51 \cdot 10^{-3}$ for the individual straight lines. Evaluating the criterion $F(15,17) = 1.13$, $F_{0.975} = 2.72$ one can conclude that the model is valid in comparison with the model of three independent straight lines. Of interest is the contribution of the mesomeric effect (as δ in Eq. (6)) for substituents in Classes 2 and 3. The parallel with the change in the reaction constant with distance suggests that the +M effect is related to the ability of the first atom in the substituent (counted from the site of bonding to the aromatic ring) to provide a lone electron pair whereas the −M effect is related to the ability of the second atom to accept the lone electron pair.

Using Eq. (6), the Gibbs energy of a process in derivatives substituted in the *meta* and *para* positions can be written as

$$\Delta G_m \;=\; \Delta G_{iso}^0 \;+\; \rho_{iso}(t_m - t_m^0) \tag{7}$$

$$\Delta G_p \;=\; \Delta G_{iso}^0 \;+\; \rho_{iso}[t_p^0 \;+\; (I + \delta\Delta M)]\,(t_m - t_m^0)\,, \tag{8}$$

where ΔG_{iso}^0 is the Gibbs energy corresponding to the substitution derivative for which $t_m = t_m^0$ (the standard state is determined by the point of intersection of the straight lines in Fig. 1 rather than by the unsubstituted compound). The constants I and $\delta\Delta M$ are so adjusted that the reaction constant ρ_{iso} is the same for the *meta* and *para* substitution, consistent with the Hammett equation. Write the relation between the Hammett substituent constants σ_m and the t_m value in the form

$$\sigma = \alpha \, [t_{\mathrm{m}} - t_{\mathrm{H}}] = \alpha \, [(t_{\mathrm{m}} - t_{\mathrm{m}}^0) - (t_{\mathrm{H}} - t_{\mathrm{m}}^0)] \, . \tag{9}$$

Rearranging Eq. (9) and inserting the term $(t_{\mathrm{H}} - t_{\mathrm{m}})$ in Eqs (7) and (8) we obtain the equations

$$\Delta G_{\mathrm{m}} = \Delta G_{\mathrm{iso}}^0 + \rho_{\mathrm{iso}} \, (t_{\mathrm{H}} - t_{\mathrm{m}}^0) + \rho_{\mathrm{iso}} \, \sigma/\alpha = \Delta G^0 + \rho \, \sigma_{\mathrm{m}} \tag{10}$$

and

$$\Delta G_{\mathrm{p}} = \Delta G_{\mathrm{iso}}^0 + \rho_{\mathrm{iso}} \, [t_{\mathrm{p}}^0 + (I + \delta\Delta M)] \, (t_{\mathrm{H}} - t_{\mathrm{m}}^0) + \rho_{\mathrm{iso}} \, [t_{\mathrm{p}}^0 + (I + \delta\Delta M)] \, \sigma/\alpha =$$

$$= \Delta G^0 + \rho \, [t_{\mathrm{p}}^0 + (I + \delta\Delta M)] \, \sigma_{\mathrm{m}} = \Delta G^0 + \rho \, \sigma_{\mathrm{p}} \, . \tag{11}$$

Constant I accounts for the change in the inductive effect on replacing the substituents in the *meta* and *para* positions, whereas the term $\delta\Delta M$ accounts for the additional substituent effect from the *para* position due to the mesomeric effect. In this sense, Eqs (10) and (11) are forms of the Hammett equation (1), whose validity for the set studied has been verified.

Analysis by the PLS Method

The PLS method was applied to the relationship between the sets of *meta* substituted derivatives (omitting experiments No. 26, 29 and 37 in Table II) and *para* substituted derivatives. If the set of *meta* derivatives constituted the independent (explanatory) matrix, the first latent variable described 97.23% variability of the data and, at the same time, 89.62% variability of the *para* derivative matrix. Comparison with the results given in Table IV demonstrates that in the case of the *meta* derivative matrix, nearly all the explainable variability by one latent variable is involved. This is not true of the variability of the *para* derivative matrix where the difference of approximately 8% is due to additional substituent effects from the *para* position (mesomeric effect). In view of the extent of explained variability of the *meta* matrix, which is comparable to the experimental error, there is no point in extending the interpretation with another latent variable. Roughly the same conclusions are arrived at if the explanatory matrix of *para* substituted derivatives (first latent variable explains 97.82% variability) for the matrix of *meta* derivatives (88.18%) is considered.

REFERENCES

1. Hammett L. P.: *Physical Organic Chemistry. Reaction Rates, Equilibria, and Mechanisms.* McGraw-Hill, New York 1970.
2. Wold S.: Chem. Scr. *5*, 97 (1974).
3. Wold S., Sjostrom M.: Chem. Scr. *2*, 49 (1972).
4. Sjostrom M., Wold S.: Chem. Scr. *6*, 114 (1974).
5. Sjostrom M., Wold S.: Chem. Scr. *9*, 200 (1976).
6. Wold S., Sjostrom M. in: *Correlation Analysis in Chemistry. Recent Advances* (J. Shorter and N. B. Chapman, Eds). Plenum Press, New York 1978.
7. Exner O.: *Correlation Analysis of Chemical Data*. Plenum Press, New York 1988.
8. Exner O.: Prog. Phys. Org. Chem. *18*, 129 (1990).
9. Exner O. in: *Correlation Analysis in Chemistry. Recent Advances* (J. Shorter and N. B. Chapman, Eds). Plenum Press, New York 1978.
10. Shorter J.: *Correlation Analysis of Organic Reactivity*. Wiley, New York 1982.
11. Hansch C., Leo A., Taft R. W.: Chem. Rev. *91*, 165 (1991).
12. Zalewski R. in: *Similarity Models in Organic Chemistry, Biochemistry and Related Fields* (R. I. Zalewski, T. M. Krygowski and J. Shorter, Eds), p. 455. Elsevier, Amsterdam 1991.
13. Sjostrom M., Wold S.: Acta Chem. Scand., B *35*, 537 (1981).
14. Alunni S., Clementi S., Edlund U., Johnels D., Hellberg S., Sjostrom M., Wold S.: Acta Chem. Scand., B *37*, 47 (1983).
15. Alunni S., Clementi S., Ebert C., Linda P., Musumarra G., Edlund V., Sjostrom M., Wold S.: Gazz. Chim. Ital. *116*, 679 (1986).
16. Nieuwdorp G. H. E., De Ligny C. L., van Houwelingen H. C.: J. Chem. Soc., Perkin Trans. 2 *1979*, 537.
17. Hnyk D.: Collect. Czech. Chem. Commun. *55*, 55 (1990).
18. Haldna U. L., Juga R. J., Tuulmets A. V., Juriado T. J.: J. Chem. Soc., Perkin Trans. 2 *1987*, 1559.
19. Sergeev N. M., Tschertkov V. A.: Dokl. Akad. Nauk SSSR *286*, 1186 (1986).
20. Livingstone D. J., Evans D. A., Saunders M. R.: J. Chem. Soc., Perkin Trans. 2 *1992*, 1545.
21. Holik M., Matejkova B.: Collect. Czech. Chem. Commun. *55*, 261 (1990).
22. Exner O., Budesinsky M.: Magn. Reson. Chem. *27*, 27 (1989).
23. Mussumarra G., Scarlata G., Wold S.: Gazz. Chim. Ital. *111*, 499 (1981).
24. Reynolds W. F., Gomes A., Maron A., McIntyre D. W., Maunder R. G., Tanin A., Wong H. E., Hamer G. K., Peat I. R.: Can. J. Chem. *61*, 2367 (1983).
25. Azzaro M., Geribaldi S., Videau B., Chastrette M.: Org. Magn. Reson. *22*, 11 (1984).
26. Hutton H. M., Kunz K. R., Bozek J. D., Blackburn B. J.: Can. J. Chem. *65*, 1316 (1987).
27. Johnels D., Edlund U., Johansson E., Coschi M.: J. Chem. Soc., Perkin Trans. 2 *1989*, 1773.
28. Pytela O., Jirman J., Lycka A.: Collect. Czech. Chem. Commun. *54*, 2399 (1989).
29. Zalewski R. I.: J. Chem. Soc., Perkin Trans. 2 *1986*, 495.
30. Ludwig M., Wold S., Exner O.: Acta Chem. Scand. *46*, 549 (1992).
31. Swain C. G., Lupton E. C.: J. Am. Chem. Soc. 90, 4328 (1968).
32. Clementi S., Fringuelli F., Linda P., Savelli G.: Gazz. Chim. Ital. *105*, 281 (1975).
33. Clementi S., Fringuelli F.: Anal. Chim. Acta *103*, 477 (1978).
34. Kazakov V. P., Korechkov Yu. D.: Zh. Fiz. Khim. *53*, 89 (1979).
35. Weeks G., Horak V.: J. Org. Chem. *45*, 2068 (1980).
36. Swain C. G., Unger S. H., Rosenquist N. R., Swain M. S.: J. Am. Chem. Soc. *105*, 492 (1983).
37. Reynolds W. F., Topsom R. D.: J. Org. Chem. *49*, 1989 (1984).

38. Charton M.: J. Org. Chem. *49*, 1997 (1984).
39. Spajner M. C., De Ligny C. L.: J. Chem. Res., Synop. *1986*, 176.
40. Taft R. W., Topsom R. D.: Prog. Phys. Org. Chem. *16*, 1 (1987).
41. Topsom R. D.: Prog. Phys. Org. Chem. *16*, 125 (1987).
42. Topsom R. D.: Prog. Phys. Org. Chem. *16*, 193 (1987).
43. Topsom R. D.: Prog. Phys. Org. Chem. *16*, 287 (1987).
44. Vereshagin A. N.: *Induktivnyi efekt*. Nauka, Moscow 1987.
45. Taft R. W., Abboud J. L. M., Anvia F., Berthelot M., Fujio M., Gal J.-F., Headley A. D., Henderson W. G., Koppel I., Qian J. H., Moshima M., Taagepera M., Ueji S.: J. Am. Chem. Soc. *110*, 1797 (1988).
46. Cook I. B., Sadek M., Ternai B.: Aust. J. Chem. *42*, 259 (1989).
47. Cook I. B.: Aust. J. Chem. *42*, 1493 (1989).
48. Pross A., Radom L., Taft R. W.: J. Org. Chem. *45*, 818 (1980).
49. Gastaiger J., Hutchings M. G.: J. Chem. Soc., Perkin Trans. 2 *1984*, 559.
50. Masuda S., Inamoto N.: Tetrahedron Lett. *26*, 6469 (1985).
51. Marriott S., Silvestro A., Topsom R. D.: J. Chem. Soc., Perkin Trans. 2 *1988*, 457.
52. Fadhil G. F., Godfrey M.: J. Chem. Soc., Perkin Trans. 2 *1988*, 133.
53. Godfrey M.: J. Chem. Soc., Perkin Trans. 2 *1989*, 1681.
54. Taft R. W., Price E., Fox I. R., Lewis I. C., Andersen K. K., Davis G. T.: J. Am. Chem. Soc. *85*, 709 (1963).
55. Taft R. W., Price E., Fox I. R., Lewis I. C., Andersen K. K., Davis G. T.: J. Am. Chem. Soc. *85*, 3146 (1963).
56. Happer D. A. R.: Aust. J. Chem. *35*, 21 (1982).
57. Pytela O., Ludwig M., Vecera M.: Collect. Czech. Chem. Commun. *51*, 2143 (1986).
58. Hoefnagel A. J., Wepster B. M.: J. Chem. Soc., Perkin Trans. 2 *1989*, 977.
59. Unger S. H., Hansch C.: Prog. Phys. Org. Chem. *12*, 91 (1976).
60. Gallo R.: Prog. Phys. Org. Chem. *14*, 115 (1983).
61. Ebert C., Linda P., Alunni S., Clementi S., Cruciani G., Santini S.: Gazz. Chim. Ital. *120*, 29 (1990).
62. Shorter J. in: *Similarity Models in Organic Chemistry, Biochemistry and Related Fields* (R. I. Zalewski, T. M. Krygowski and J. Shorter, Eds), p. 77. Elsevier, Amsterdam 1991.
63. Ehrenson S. R., Brownlee T. C., Taft R. W.: Prog. Phys. Org. Chem. *10*, 1 (1973).
64. Istomin B. I., Pivovarov S. A., Selivanov V. F., Gidaspov B. V., Istomina S. N.: Org. React. *12*, 287 (1975).
65. Exner O.: Collect. Czech. Chem. Commun. *41*, 1516 (1976).
66. Exner O.: Org. React. *14*, 353 (1977).
67. Koppel I. A., Molder U. H.: Org. React. *21*, 234 (1984).
68. Holik M.: Magn. Reson. Chem. *30*, 189 (1992).
69. Ludwig M., Baron V., Kalfus K., Pytela O., Vecera M.: Collect. Czech. Chem. Commun. *51*, 2135 (1986).
70. Exner O., Kalfus K.: Collect. Czech. Chem. Commun. *41*, 569 (1976).
71. Hoefnagel A. J., Wepster B. M.: Collect. Czech. Chem. Commun. *55*, 119 (1990).
72. Exner O., Bocek K.: Collect. Czech. Chem. Commun. *38*, 50 (1973).
73. Laurence C., Berthelot M.: J. Chem. Soc., Perkin Trans. 2 *1979*, 98.
74. Bohm S., Kuthan J.: Int. J. Quant. Chem. *26*, 21 (1984).
75. Gilliom R. D., Beck J.-P., Purcell W. P.: J. Comput. Chem. *6*, 437 (1985).
76. La Manna G., Tschinke V., Paoloni L.: J. Chem. Soc., Perkin Trans. 2 *1985*, 1393.
77. Sotomatsu T., Murata Y., Fujita T.: J. Comput. Chem. *10*, 94 (1989).

78. McMahon T. B., Kebarle P.: J. Am. Chem. Soc. *99*, 2222 (1977).

79. Dippy J. F. J., Hughes S. R. C., Kitchiner B. C.: J. Chem. Soc. *1964*, 1275.

80. Wilson J. M., Gore N. E., Sawbridge J. E., Cardenas-Cruz F.: J. Chem. Soc., B *1967*, 852.

81. Matsui T., Ko H. C., Hepler L. G.: Can. J. Chem. *52*, 2906 (1974).

82. Steigman J., Sussman D.: J. Am. Chem. Soc. *89*, 6406 (1967).

83. Weiner P. H.: J. Am. Chem. Soc. *95*, 5845 (1973).

84. Hoefnagel A. J., Hoefnagel M. A., Wepster B. M.: J. Org. Chem. *43*, 4720 (1978).

85. Hoefnagel A. J., Oosterbeek W., Wepster B. M.: J. Org. Chem. *49*, 1993 (1984).

86. Kalfus K., Kroupa J., Vecera M., Exner O.: Collect. Czech. Chem. Commun. *40*, 3009 (1975).

87. Van Meurs F., Hoefnagel A. J., Wepster B. M., Van Bekkum H.: J. Organomet. Chem. *142*, 299 (1977).

88. Exner O., Lakomy J.: Collect. Czech. Chem. Commun. *35*, 1371 (1970).

89. Jahagirdar D. V., Shelke D. N., Deshpande R. G.: J. Chem. Soc., Perkin Trans. 2 *1977*, 1513.

90. Kalfus K., Vecera M., Exner O.: Collect. Czech. Chem. Commun. *35*, 1195 (1970).

91. Kalfus K., Vecera M.: Collect. Czech. Chem. Commun. *37*, 3607 (1972).

92. Kozachenko A. G., Matrosov E. I., Kabachnik M. I.: Izv. Akad. Nauk SSSR, Ser. Khim. *1976*, 2440.

93. Pytela O.: Collect. Czech. Chem. Commun. *55*, 42 (1990).

Translated by P. Adamek.