

## CHEMOMETRIC ANALYSIS OF SUBSTITUENT EFFECTS. IX. ALTERNATIVE INTERPRETATION OF SUBSTITUENT EFFECTS (AISE) – ORTHOGONAL MODEL

Oldrich PYTELA

*Department of Organic Chemistry, University of Pardubice, 532 10 Pardubice, Czech Republic;  
e-mail: pytela@hlab.upce.cz*

Received December 5, 1995

Accepted January 12, 1996

Continuing the previous communications, the present paper suggests an orthogonal model of alternative interpretation of substituent effects (AISE). Following this approach, substituents are classified into three classes. Those of class II are defined as substituents whose atom bound to the basic skeleton bears a character of nucleophile, i.e. it is able of an intramolecular nucleophilic interaction with the reaction centre. The class III includes substituents whose atom bound to the basic skeleton is electrophilic in nature and can correspondingly interact with the reaction centre too. The class I includes the substituents possessing none of the above-mentioned characteristics. Within the model suggested the substituent effects can be described by a family of three lines corresponding to the substituent classes described with a single substituent constant at the coordinate axis. The validity of the model suggested has successfully been tested by comparison with other correlation equations using 318 reaction series taken from literature. It has been found that the value of the point of intersection at the coordinate axis (the isosubstituent constant) is smaller in processes with electron deficit and greater in those with electron excess in the reaction centre. The slopes of the individual lines in the family of lines (the reaction constants) decrease in the order  $II > I > III$  in most cases, the only exception being the processes with direct conjugation between a negatively charged reaction centre and the substituent and some gas phase processes.

**Key words:** Chemometrics; Substituent effects; Alternative interpretation of substituent effects.

The quantitative approach to description of substituent effects has passed through a certain development<sup>1-9</sup> since the first formulation by Hammett, including changes in the basic principles. Hammett's approach<sup>1</sup> can be denoted as formal in the context given. The substituent constants  $\sigma_m$  and  $\sigma_p$  involve not only the properties of substituent but also its position in the benzene nucleus<sup>10</sup>. The advantage of this approach is in its having a single reaction constant, its drawback consists in the necessity of parametrization of the substituent constants for other basic skeletons or other types of reaction centres<sup>11-14</sup>. The second principle of quantitative description of substituent effects is based on separate description of elementary substituent effects defined in a certain way<sup>2-4,6,7,15-24</sup>. This approach is clearly more general since in an ideal case the substituent constants do not involve the properties of basic skeleton. The fundamental pro-

blem of this approach consists in the definition of the elementary substituent effects and their separation. In a previous paper<sup>25</sup> we published a qualitatively new principle of description of substituent effects denoted as alternative interpretation of substituent effects (AISE). The central dogma of this approach rests on the presumption of a single property of substituent described by a single substituent constant. This property is transferred to the reaction centre by three different ways<sup>26,27</sup> depending on the interaction type in the triad: reaction centre–basic skeleton–substituent, which results in the classification of substituents into three classes. If a substituent has no  $\pi$  electrons at the atom connecting it to the basic skeleton, it is classified as the class I substituent. The substituents of the class II have a free electron pair at the said first atom, whereas those of the class III contain a multiple bond – polarized in the direction from the basic skeleton – between the first and the second atoms. The effects of the class I substituents are described by the substituent constant  $\sigma^i$  (refs<sup>25,28</sup>) identical with  $\sigma_I$  constant. Substituents of classes II and III exhibit additional effects proportional to the same substituent constant. Mathematically, the description of substituent effects in the formerly published<sup>25</sup> concept is expressed by a family of three straight lines with a single interpreting variable  $\sigma^i$  according to Eq. (1)

$$\log k = \log k_0 + \rho_I \mathbf{1}_I(\sigma^i - \sigma_0^i) + \rho_X \mathbf{1}_X(\sigma^i - \sigma_0^i) + \rho_{XY} \mathbf{1}_{XY}(\sigma^i - \sigma_0^i) , \quad (1)$$

where  $\log k$  is logarithm of rate or equilibrium constant,  $\log k_0$  is the same quantity at the point of intersection of the straight lines,  $\rho_I$ ,  $\rho_X$ , and  $\rho_{XY}$  represent the reaction constants related to the substituents of the classes I, II, and III, respectively, and the terms  $\mathbf{1}_I$ ,  $\mathbf{1}_X$ , and  $\mathbf{1}_{XY}$  are the multiplying constants which assume the values of 1 or 0 depending on whether or not the substituent involves the given type of interaction, respectively. The point of intersection of the straight lines, which was denoted as the isoeffect point  $\sigma_0^i$ , has been declared to represent a universal constant with the calculated value of 0.534 (ref.<sup>25</sup>). The validity of AISE principle was verified<sup>25</sup> with good results using 318 series of experimental data taken from literature, covering substituent effects in a wide variety of chemical models (acid-base equilibria of substituted aliphatic, alicyclic and arylaliphatic carboxylic acids, benzoic acids, naphthalenecarboxylic acids, benzenesulfonamides, anilines, phenols, quinuclidines, pyridines and other heterocycles, and moreover solvolyses of alicyclic arenesulfonates and reactions involving formation of carbocations).

Equation (1) also allows an interpretation different from that published in the previous paper<sup>25</sup>. The reaction constants  $\rho_X$  and  $\rho_{XY}$  can be interpreted not as constants expressing an additional effect of substituent due to interactions mediated by  $\pi$  electrons but as ones reflecting an effect of a certain type of substituent as a whole. A model defined in this way can be denoted as orthogonal. Also the isoeffect point  $\sigma_0^i$

need not necessarily be a general constant but can differ in different sets depending on the type and conditions of the given process. The aim of the present paper is to verify whether the AISE modifications given result in a more suitable model for description of substituent effects.

## THEORETICAL

The concept of alternative interpretation of substituent effects (AISE) formulated in the previous paper<sup>25</sup> starts from the idea of a single basic effect transmitted through the  $\sigma$  skeleton (with the reaction constant  $\rho_I$ ) and additional effects transmitted due to  $\pi$  electron interaction between substituent and basic skeleton (with the reaction constants  $\rho_X$  and  $\rho_{XY}$ ). This model can be denoted as non-orthogonal with regard to the possible dependence between the interpreting variables in Eq. (1). The above-declared orthogonal model is more general. Also in this approach, substituents are divided into three classes according to the type of their interaction with the basic skeleton. The substituents of class II are defined as ones in which the atom next to the basic skeleton is of nucleophilic nature and is able of an intramolecular nucleophilic interaction with the reaction centre. Similarly, class III involves substituents whose first atom is of electrophilic nature and able of corresponding interaction with the reaction centre. Class I substituents exhibit none of the above-mentioned properties. The most common substituents arranged according to these criteria are presented in Table I. From the point of view of the approach described it makes no difference whether or not the electrons of the key atom of substituent are involved in conjugation within the substituent. The extent of the intra-substituent interaction is included in the value of substituent constant  $\sigma^i$ , which can be seen from a comparison of suitable substituents in Table I. Regardless of chemical structure of substituent, e.g., the  $\text{NHCOCH}_3$  group belongs to class II, the  $\text{CONH}_2$  and  $\text{COO}^-$  groups to class III, and the  $\text{NH}_3^+$  group to class I etc. An unequivocal classification of a substituent in one of the three classes declared allows us to define an orthogonal AISE model by Eq. (2)

$$\log k = \log k_0 + \rho_I \delta_I (\sigma^i - \sigma_0^i) + \rho_N \delta_N (\sigma^i - \sigma_0^i) + \rho_E \delta_E (\sigma^i - \sigma_0^i), \quad (2)$$

which is formally identical with Eq. (1). The orthogonality of model (2) is given by the multiplying constants  $\delta_I$ ,  $\delta_N$  and  $\delta_E$  of the type of Kronecker's delta which assume the values 1 or 0 depending on whether or not the substituent belongs to the given class, respectively. The respective reaction constants in Eq. (2) express the sensitivity to the substituents of the first ( $\rho_I$ ), second ( $\rho_N$ ), or third ( $\rho_E$ ) class, and they can generally show various mutual relations of magnitude. The straight lines characterized by these reaction constants intersect at a single point with the value of substituent constant equal

to  $\sigma_0^i$ . In contrast to the previous communication<sup>25</sup> we presume the  $\sigma_0^i$  constant not to be universal and identical for all processes.

The introduction of orthogonal model (2) is advantageous also from the standpoint of mathematical-statistical treatment because no multicollinearity can arise between the explanatory variables. For a practical application of the AISE orthogonal model the minimum set of substituents should – in accordance with the data of Table I – involve methyl and trifluoromethyl substituents (class I), amino (and/or *N*-methylamino and/or *N,N*-dimethylamino), methoxy and some of halogen substituents (class II) and furthermore acetyl and nitro substituents (as the best ones of class III). Unfortunately, it must be stated that this requirement of suitable representation of substituents is met by only a small part of data available in literature.

TABLE I  
Substituent constants  $\sigma^i$  used in AISE models<sup>25,28</sup>

Class I <sup>a</sup>		Class II <sup>b</sup>		Class III <sup>c</sup>	
Substituent	$\sigma^i$	Substituent	$\sigma^i$	Substituent	$\sigma^i$
H	0.000	NH <sub>2</sub>	0.089	CHO	0.385
Me	-0.040	NHCH <sub>3</sub>	0.094	COCH <sub>3</sub>	0.286
Et	-0.048	N(CH <sub>3</sub> ) <sub>2</sub>	0.089	COOH	0.264
Pr	-0.054	NHCOCH <sub>3</sub>	0.230	COOCH <sub>3</sub>	0.274
Bu	-0.058	OH	0.157	COOC <sub>2</sub> H <sub>5</sub>	0.265
iPr	-0.062	OCH <sub>3</sub>	0.220	CONH <sub>2</sub>	0.250
iBu	-0.056	OC <sub>6</sub> H <sub>5</sub>	0.278	CN	0.525
<i>sec</i> -Bu	-0.068	OCOCH <sub>3</sub>	0.290	NO <sub>2</sub>	0.606
<i>tert</i> -Bu	-0.084	SH	0.239	SO <sub>2</sub> CH <sub>3</sub>	0.551
neoPe	-0.064	SCH <sub>3</sub>	0.217	SO <sub>2</sub> NH <sub>2</sub>	0.423
CycloHx	-0.061	F	0.343		
Benzyl	-0.018	Cl	0.374		
Vinyl	0.041	Br	0.384		
Ethynyl	0.246	I	0.353		
Phenyl	0.078				
CF <sub>3</sub>	0.372				

<sup>a</sup>  $\delta_I = 1, \delta_N = \delta_E = 0$ ; <sup>b</sup>  $\delta_N = 1, \delta_I = \delta_E = 0$ ; <sup>c</sup>  $\delta_E = 1, \delta_I = \delta_N = 0$ .

## CALCULATIONS

For verification of validity of the orthogonal AISE model we used a set of 318 reaction series of experimental data taken from literature, describing the substituent effect on a wide variety of chemical models (acid-base equilibria of substituted aliphatic, alicyclic and arylaliphatic carboxylic acids, benzoic acids, naphthalenecarboxylic acids, benzenesulfonamides, anilines, phenols, quinuclidines, pyridines and other heterocycles, and moreover solvolyses of alicyclic arenesulfonates and reactions involving formation of carbocations). For the calculations were used only series with at least 6 out of 32 representative substituents. A detailed specification of the set used for the test was given in ref.<sup>25</sup>.

The calculation of the isosubstituent constant  $\sigma_0^i$  and respective reaction constants of Eq. (2) made use of a combination of one-dimensional optimization and triple linear regression in the sense of the least squares treatment, which is an analogy of evaluation of the isokinetic temperature<sup>29</sup>. The objective function used was constructed as a sum of squares of differences between experimental and calculated values of  $\log k$  in Eq. (2) for the given experimental series. The calculations were carried out with the use of our own programs.

## RESULTS AND DISCUSSION

### *Evaluation of Orthogonal AISE Model Based on Interpreted Variability of Experimental Data*

For comparing the validity of orthogonal AISE model we selected three other correlation equations. The first of them is represented by the known dual model with the substituent constants  $\sigma_I$  (inductive effect),  $\sigma_R$  (mesomeric effect) using the parametrization according to ref.<sup>8</sup>. The second reference relation was the correlation equation suggested by Charton<sup>7</sup>, with the substituent constants  $\sigma_1$  (localized effect),  $\sigma_d$  (delocalized effect), and  $\sigma_e$  (additional effect of the substituent affected by the reaction centre). The third relation being compared was the AISE model described by Eq. (1) (ref.<sup>25</sup>) with the parametrization according to Table I. The same parametrization was also used for Eq. (2). The explained variability obtained by applying all the three reference models and the orthogonal AISE model on the set of experimental data described in ref.<sup>25</sup> is given in Table II in a structured way according to the type of the chemical compound. If we at first compare both the AISE models, it is obvious that the orthogonal model (2) is unequivocally more successful than model (1). The most distinct improvement can be seen in the processes with possible interactions between the reaction centre and substituent: typical examples are dissociations of *para* substituted phenols or anilines and formation of carbocations. This is obviously caused by the distinctly different value of the isosubstituent constant  $\sigma_0^i$  for the individual processes as compared with the "aver-

age" value of 0.534 adjusted for the whole test set of data according to Eq. (1) (ref.<sup>25</sup>). If the explained variability of all the relations tested in Table II is compared, it can be stated that the orthogonal AISE model with a single substituent constant is better in all the cases than the dual model<sup>17</sup> with two substituent constants, being also better – but for two exceptions (dissociations of phenols and anilines) – than the Charton relation<sup>7</sup> with three substituent constants.

TABLE II

Comparison of interpreted variability ( $V$ , %) and number of degrees of freedom ( $v$ , in brackets) of standardized data according to the type of process and correlation equation

No.	Compounds	V, % (v)							
		$\sigma_I, \sigma_R$ (ref. <sup>8</sup> )		$\sigma_I, \sigma_d, \sigma_e$ (ref. <sup>7</sup> )		Eq. (1) <sup>a</sup> (ref. <sup>25</sup> )		Eq.(2) <sup>a</sup>	
1	Aliphatic acids	97.74	(73)	97.24	(55)	97.76	(46)	98.62	(65)
2	Arylaliphatic carboxylic acids	93.32	(44)	96.70	(27)	95.31	(29)	97.43	(52)
3	Benzoic acids	93.84	(664)	93.81	(625)	95.52	(623)	96.42	(545)
3A	<i>meta</i> substituted	95.07	(342)	94.18	(313)	95.78	(312)	97.00	(304)
3B	<i>para</i> substituted	92.74	(269)	93.38	(235)	95.25	(234)	95.74	(241)
4	Naphthalenecarboxylic acids	95.77	(28)	94.31	(15)	95.31	(15)	96.08	(36)
5	Benzenesulfonamides	98.01	(115)	98.26	(106)	97.95	(106)	99.45	(64)
6	Alicyclic arenesulfonates	93.22	(51)	94.26	(27)	93.02	(28)	97.19	(49)
7	Carbocations	94.22	(74)	96.82	(47)	94.75	(56)	97.30	(70)
7B	<i>para</i> substituted	95.50	(41)	97.78	(19)	95.40	(26)	97.93	(47)
8	Anilines	95.35	(121)	98.81	(82)	95.73	(100)	97.47	(106)
8B	<i>para</i> substituted	96.94	(41)	99.03	(18)	94.53	(27)	97.01	(46)
9	Phenols	95.02	(171)	97.87	(133)	94.11	(148)	97.75	(150)
9B	<i>para</i> substituted	94.66	(78)	98.85	(44)	91.79	(56)	97.38	(89)
10	Quinuclidines	97.97	(52)	97.04	(27)	97.54	(22)	98.28	(76)
11	Pyridines	89.79	(133)	93.59	(97)	93.89	(106)	94.67	(146)
12	Other heterocycles	94.21	(83)	95.81	(57)	94.95	(58)	97.14	(79)
13	Other compounds	94.59	(79)	95.56	(53)	96.34	(55)	97.20	(82)
14	All the compounds	94.38	(2 157)	95.64	(2 039)	95.36	(2 085)	97.06	(1 524)

<sup>a</sup> For parametrization see Table I.

*Comparison of Values of Isosubstituent Constant  $\sigma_0^i$  for Individual Types of Processes*

The treatment of experimental data taken from literature, as specified in Calculations, with the use of Eq. (2) gave the relation of frequency on the value of the isoeffect constant  $\sigma_0^i$  which is summarized in Table III. From the table it follows that most values (77%) are smaller than 1.5, the rest being greater than this value. The frequency distribution in the first group is unimodal and almost symmetrical about the interval of values 0.4–0.5. An overwhelming majority of the values of isoeffect constant in the second group are of the  $10^3$  order of magnitude, which can be interpreted as the impossibility of determination of point of intersection of the straight lines in model (2) for the given series of data within experimental accuracy. A more detailed analysis shows that this mostly concerns the data involving insufficient numbers (less than 3) of substituents of class III. The magnitude of value of isoeffect constant depends on the type of process. Lower values are characteristic of processes with electron deficit at the reaction centre such as in the solvolyses of alicyclic arylsulfonates, formation of carbocations, or protonation of pyridines. Medium values were found with the processes involving substrates with aromatic nuclei without any possibility of conjugation between the reaction centre and substituent. Typical examples involve dissociations of benzoic acids and benzenesulfonamides. High values of isosubstituent constant are characteristic of processes with an electron excess at the reaction centre combined with the possibility of direct conjugation of the reaction centre with the substituents of class III, as it is the case with dissociations of phenol and – partially – protonations of anilines. If there are – at these conditions – few substituents of the class III among the data, it is impossible to determine the point of intersection of the straight lines. The

TABLE III

Dependence of occurrence frequency upon value of isoeffect constant  $\sigma_0^i$  according to Eq. (2) for data specified in Calculations

Interval	<i>p</i>	Interval	<i>p</i>
⟨0,0.1⟩	13	⟨0.8,0.9⟩	4
⟨0.1,0.2⟩	8	⟨0.9,1.0⟩	6
⟨0.2,0.3⟩	14	⟨1.0,1.1⟩	1
⟨0.3,0.4⟩	44	⟨1.1,1.2⟩	2
⟨0.4,0.5⟩	73	⟨1.2,1.3⟩	2
⟨0.5,0.6⟩	46	⟨1.3,1.4⟩	1
⟨0.6,0.7⟩	18	⟨1.4,1.5⟩	2
⟨0.7,0.8⟩	11	⟨1.5,10 <sup>4</sup> ⟩	73

conclusions given are of probability nature since the regions of distribution of isosubstituent constant for the said process types overlap.

*Comparison of Values of Reaction Constants According to Substituent and Process Types*

The values of reaction constants  $\rho_I$ ,  $\rho_N$ , and  $\rho_E$  obtained by treating the data specified in Calculations according to Eq. (2) varied within the interval of values usual for similar correlation relations inclusive of the signs. During the calculation, the reaction constants for each series were tested with respect to collinearity on the basis of the hypothesis of agreement between the residual spreads for model (2) and model with a single common reaction constant ( $\rho_I = \rho_N = \rho_E$ ). The processes classified as collinear at the significance level  $\alpha = 0.05$  first of all involved those on substrates with aliphatic skeleton, such as protonation of quinuclidines (88.9% of cases), dissociation of aliphatic acids (82.4%), or solvolysis of alicyclic arenesulfonates (80.0%). This agrees with the idea of negligible electrophilic or nucleophilic interactions between substituent and reaction centre mediated by the basic skeleton in these compounds. Distinctly less statistically significant collinearity was found in the processes involving direct conjugation between reaction centre and substituent, such as dissociation of *para* substituted phenols (16.7% of cases), anilines (28.6%), or formation of carbocations with *para* substituents (35.7%). The collinearity in these compounds is due to missing substituents of class III in most cases.

The magnitudes of reaction constants follow the order  $\rho_N > \rho_I > \rho_E$  unless the statistical hypothesis of their equality is valid. Exceptions from this rule were only found with the processes connected with direct conjugation between negatively charged reaction centre and substituent (dissociation of *para* substituted phenols), where the reaction constant  $\rho_E$  is greater than the reaction constants  $\rho_I \approx \rho_N$ . This result could be expected and can easily be interpreted. On the other hand, the greater sensitivity to substituents of classes II and III as compared with those of class I found in gas phase processes (dissociation of *meta* and *para* substituted benzoic acids, *para* substituted phenols, 3- and 4-substituted pyridines) has no such simple interpretation. Most probably some other transmission mechanisms make themselves felt besides those operating in more polar media of solvents which are rich in intermolecular interactions.

*This work was supported by the Grant Agency of Czech Republic, Grant No. 203-94-0122.*

## REFERENCES

1. Hammett L. P.: *Physical Organic Chemistry. Reaction Rates, Equilibria, and Mechanisms*, 2nd ed. McGraw-Hill, New York 1970.



2. Chapman N. B., Shorter J. (Eds): *Advances in Linear Free Energy Relationships*. Plenum, New York 1972.
3. Shorter J., Chapman N. B. (Eds): *Correlation Analysis in Chemistry: Recent Advances*. Plenum Press, New York 1978.
4. Charton M.: *Prog. Phys. Org. Chem.* *13*, 119 (1981).
5. Shorter J.: *Correlation Analysis of Organic Reactivity*. Wiley, New York 1982.
6. Taft R. W., Topsom R. D.: *Prog. Phys. Org. Chem.* *16*, 1 (1987).
7. Charton M.: *Prog. Phys. Org. Chem.* *16*, 287 (1987).
8. Exner O.: *Correlation Analysis of Chemical Data*. Plenum, New York 1988.
9. Zalewski R. I., Krygowski T. M., Shorter J. (Eds): *Similarity Models in Organic Chemistry, Biochemistry and Related Fields*. Elsevier, Amsterdam 1991.
10. Shorter J.: *Pure Appl. Chem.* *66*, 2452 (1994).
11. Hammett L. P.: *J. Am. Chem. Soc.* *59*, 96 (1937).
12. Brown H. C., Okamoto Y.: *J. Am. Chem. Soc.* *80*, 4979 (1958).
13. Yukawa Y., Tsuno Y.: *Bull. Chem. Soc. Jpn.* *32*, 971 (1959).
14. Yoshioka M., Hamamoto K., Kubota T.: *Bull. Chem. Soc. Jpn.* *35*, 1723 (1962).
15. Taft R. W.: *Steric Effects in Organic Chemistry* (M. S. Newman, Ed.). Wiley, New York 1956.
16. Swain C. G., Lupton E. C.: *J. Am. Chem. Soc.* *90*, 4328 (1968).
17. Ehrenson S., Brownlee R. T. C., Taft R. W.: *Prog. Phys. Org. Chem.* *10*, 1 (1973).
18. Nieuwdorp G. H. E., de Ligny C. L., van Houwelingen H.: *J. Chem. Soc., Perkin Trans. 2* *1979*, 537.
19. Swain C. G., Unger S. H., Rosenquist N. R., Swain M. S.: *J. Am. Chem. Soc.* *105*, 492 (1983).
20. Hansch C., Leo A., Taft R. W.: *Chem. Rev.* *91*, 165 (1991).
21. Drago R. S., Dadmun A. P.: *J. Am. Chem. Soc.* *115*, 8592 (1993).
22. Exner O., Friedl Z.: *Prog. Phys. Org. Chem.* *19*, 259 (1993).
23. Bowden K., Grubbs E. J.: *Prog. Phys. Org. Chem.* *19*, 183 (1993).
24. Drago R. S., Dadmun A. P.: *J. Am. Chem. Soc.* *116*, 1792 (1994).
25. Pytela O.: *Collect. Czech. Chem. Commun.* *60*, 1502 (1995).
26. Pytela O.: *Collect. Czech. Chem. Commun.* *59*, 159 (1994).
27. Pytela O.: *Collect. Czech. Chem. Commun.* *59*, 381 (1994).
28. Pytela O., Halama A.: *Collect. Czech. Chem. Commun.* *60*, 1316 (1995).
29. Pytela O., Vecera M., Vetesnik P.: *Collect. Czech. Chem. Commun.* *46*, 898 (1981).