Modeling of substituent effects: self-parametrization, a case of interdependence of the coefficients, illustrated on the linear plus cross-term models

Bruno Blaive [a] and Jean-Claude Ziegler [b]

[a] *CNRS, École Nat. Sup. Chimie, Faculty of St.-Jerome, F-13397 Marseille cedex 20, France*
E-mail: blaive@spi-chim.u-3mrs.fr
[b] *Centre du Medicament, 30 rue Lionnois, F-54000 Nancy, France*

The effect of one or several substituents X_i on a molecular property f (reaction rate, spectroscopic property, etc.) is often described in terms of substituent parameters $\lambda_i(X)$ depending on the nature X of the substituent, and the position i of substitution. We call *self-parametrization* the case when the substituents are parameterized by the value $\lambda_i(X) = f_i(X)$ of f itself in the molecule monosubstituted by X in position i. On the example of the linear plus cross-term models, we show that self-parametrization implies that the coefficients of the model are interdependent. The theoretical relations between them are given, for equivalent or nonequivalent positions of substitution, and tested on measured acid–base equilibrium constants, and NMR coupling constants.

KEY WORDS: molecular property, substituents, self-parameterization

1.   Introduction

Most of the properties which can be measured on molecules have been studied in series of compounds, as functions of the nature of the substituents X_i that are present in one or several positions i. These properties are chemical equilibrium constants, reaction rates, NMR chemical shifts or coupling constants, bond vibration frequencies, and others [1–6]. In many cases, the effect of a substituent X on a property f can be represented by a scalar parameter $\lambda_i(X)$ (or several $\lambda_i^{(k)}$, i.e., a vector parameter λ_i), which depends on the position i of the substituent, and on its nature X. The function λ_i can be either a physicochemical constant, or a parameter optimized from the data (indicator variable) [7–9]. For n substituents, we have

$$f(X_1, \ldots, X_n) = F\big[\lambda_1(X_1), \ldots, \lambda_n(X_n)\big], \tag{1}$$

where F is a scalar function of n variables.

Let us denote by $f_i(X) \equiv f(H, \ldots, H, X, H, \ldots, H)$ the value of f measured in the molecule monosubstituted by X in position i, i.e., having hydrogen atoms in the

69

$n - 1$ other positions of substitution. We shall call *self-parametrization* the case when the scalar function

$$\lambda_i(X) = f_i(X) \tag{2}$$

is chosen to parameterize the substituent X in position i. In this case, equation (1) can be written as

$$f(X_1, \ldots, X_n) = F\big[f_1(X_1), \ldots, f_n(X_n)\big]. \tag{3}$$

In the present article we shall study the self-parametrization in the case when the function F is a linear plus cross-term model,

$$f(X_1, \ldots X_n) = a_0 + \sum_{i=1}^{n} a_{1i}\lambda_i(X_i) + \sum_{i=1}^{n} \sum_{j=1, j>i}^{n} a_{2ij}\lambda_i(X_i)\lambda_j(X_j) \tag{4}$$

$$= a_0 + \sum_{i=1}^{n} a_{1i} f_i(X_i) + \sum_{i=1}^{n} \sum_{j=1, j>i}^{n} a_{2ij} f_i(X_i) f_j(X_j). \tag{5}$$

Equation (4) is a simple model to be tried [10] when nonadditive substituent effects are observed on a property f. We will show that the coefficients a's of equation (5) are interdependent, and we shall verify this interdependence on several chemical examples.

To simplify the notations, the substituents have been indexed by only one index i in equations (1)–(5), agreeing that different indices i, j sometimes represent the same position of substitution. This replaces double sums $\sum_p \sum_s$ (p position of substitution, s substituent) [7] by only one sum $\sum_i$, and avoids fourfold summations in equations (4), (5).

2.   Alternative sets λ_i of substituent parametrizations compatible with equation (4)

Let $f(X_1, \ldots, X_n)$ be a scalar property depending on the substituents placed in n positions on a molecular substrate. We assume that, for each position of substitution i, there exists a parametrization $\lambda_i(X)$ such that f can be computed by equation (4).

(a)   Clearly, any set (λ_i') of substituent parametrizations such that each λ_i' is in an affine relationship with λ_i,

$$\lambda_i' = \alpha_i + \beta_i \lambda_i \quad (\beta_i \neq 0), \tag{6}$$

does not change the form of equation (4). Then, the coefficients of equation (4) are changed into

$$a_0' = a_0 - \sum_{i=1}^{n} a_{1i}\frac{\alpha_i}{\beta_i} + \sum_{i} \sum_{j, j>i} a_{2ij}\frac{\alpha_i \alpha_j}{\beta_i \beta_j}, \tag{7}$$

$$a'_{1i} = \frac{1}{\beta_i}\left(a_{1i} - \sum_{j\neq i} a_{2ij}\frac{\alpha_j}{\beta_j}\right) \quad (i = 1, \ldots, n), \tag{8}$$

$$a'_{2ij} = \frac{a_{2ij}}{\beta_i \beta_j} \qquad (i, j = 1, \ldots, n). \tag{9}$$

We notice that some of the coefficients a_0, a_{1i}, a_{2ij} could be removed, in choosing appropriate constants α_i, β_i in equation (6).

(b) Among the sets (λ'_i) of parametrizations defined by equation (6), one is the self-parametrization (f_i). Indeed, we have

$$f_i(X) = \alpha_i + \beta_i\lambda_i(X) \quad (i = 1, \ldots, n), \tag{10}$$

with the constants

$$\alpha_i = a_0 + \sum_{j,j\neq i} a_{1j}\lambda_j(H) + \sum_{j,j\neq i}\sum_{k,k>j,k\neq i} a_{2jk}\lambda_j(H)\lambda_k(H), \tag{11}$$

$$\beta_i = a_{1i} + \sum_{j,j\neq i} a_{2ij}\lambda_j(H). \tag{12}$$

(c) Reciprocally, there are no other sets (λ'_i) of parametrizations compatible with equation (4), than those defined by equation (6). In fact, if two sets (λ_i) and (λ'_i) are compatible with the form of equation (4), they satisfy equation (10) and similarly

$$f_i(X) = \alpha'_i + \beta'_i\lambda'_i(X). \tag{13}$$

This shows that one also passes from λ_i to λ'_i by an affine relationship.

(d) Finally, let us consider the particular case when the n positions of substitution are equivalent with respect to the measured property f (for example, identical positions, or symmetrical positions with respect to the active site of the molecule, when the property is a reaction rate). Then, a common substituent parametrization $\lambda_i = \lambda$ can be used for all the positions, and the coefficients $a_{1i} = a_1$ are equal, as well as the coefficients $a_{2ij} = a_2$. Equations (6) and (7)–(9) for the change of parametrization can be simply written as

$$\lambda' = \alpha + \beta\lambda, \tag{14}$$

and

$$a'_0 = a_0 - n\frac{\alpha}{\beta}a_1 + \frac{n(n-1)}{2}\frac{\alpha^2}{\beta^2}a_2, \tag{15}$$

$$a'_1 = \frac{1}{\beta}\left[a_1 - (n-1)\frac{\alpha}{\beta}a_2\right], \tag{16}$$

$$a'_2 = \frac{a_2}{\beta^2}. \tag{17}$$

There are two infinities of sets of coefficients (a_0, a_1, a_2) that are equivalent in equation (4), corresponding to the choice of α and β in equation (14). In other words, it is possible to impose the values of two parameters chosen among a_0, a_1, a_2, for example, a_0 and a_1. Therefore, there must be a combination of the three parameters which is invariant in the transformation (14). Indeed, by eliminating α and β between equations (15)–(17), we see that the quantity $4a_0 - 3a_1^2/a_2$ is invariant, i.e., we have

$$4a_0 - \frac{3a_1^2}{a_2} = 4a_0' - \frac{3a_1'^2}{a_2'}. \tag{18}$$

Note that equation (18) does not mean that the coefficients a_0, a_1, a_2 are interdependent: examples can be provided where a_0, a_1, a_2 are equal to any given number.

3.   The self-parametrization $\lambda_i(X) = f_i(X)$

(a) For the hydrogen substituent $(X = H)$ the n parametrizations f_i take the common value

$$f_i(H) = f(H, \ldots, H), \tag{19}$$

which will be denoted simply by $f(H)$. By substracting this constant from f, we define a new property

$$g(X_1, \ldots, X_n) = f(X_1, \ldots, X_n) - f(H). \tag{20}$$

In this section, we shall assume that the property f can be represented by an equation of the form (4), and, therefore, by an equation of the form (5) (see section 2). Clearly, the property g also satisfies an equation of the form (5),

$$g(X_1, \ldots, X_n) = b_0 + \sum_{i=1}^{n} b_{1i} g_i(X_i) + \sum_{i=1}^{n} \sum_{j, j>i} b_{2ij} g_i(X_i) \, g_j(X_j), \tag{21}$$

where

$$g_i(X) = g(H, \ldots, H, X, H, \ldots, H) \tag{22}$$

is the self-parametrization associated with g. The definition (20) implies

$$g(H) = g_i(H) = g(H, \ldots, H) = 0, \tag{23}$$

which shows that b_0 equals zero. In addition, equation (23) implies

$$g_i(X)[b_{1i} - 1] = 0 \tag{24}$$

for any X and i. All the b_{1i}'s are therefore equal to one, and equation (21) can be rewritten as

$$g(X_1, \ldots, X_n) = \sum_{i=1}^{n} g_i(X_i) + \sum_{i=1}^{n} \sum_{j, j>i} b_{2ij} g_i(X_i) g_j(X_j). \tag{25}$$

In terms of f, equation (25) is equivalent to equation (5) with the coefficients

$$a_0 = f(\mathrm{H})\left[1 - n + f(\mathrm{H}) \sum_{i=1}^{n} \sum_{j,j>i} a_{2ij} \right], \tag{26}$$

$$a_{1i} = 1 - f(\mathrm{H}) \sum_{j,j\neq i} a_{2ij}, \tag{27}$$

$$a_{2ij} \text{ arbitrary}. \tag{28}$$

Eliminating the constant $f(\mathrm{H})$ between the $n + 1$ equations (26), (27), we draw n relations linking the $(n^2 + n + 2)/2$ coefficients a_0, a_{1i}, a_{2ij}. These relations can be written in an infinite number of manners, since they can be combined algebraically. A simple system of n equations is obtained as follows. The n equations (27) give the $n-1$ relations

$$\frac{1 - a_{11}}{\sum_{j,j\neq 1} a_{21j}} = \frac{1 - a_{1i}}{\sum_{j,j\neq i} a_{2ij}} \quad (i = 2, \ldots, n). \tag{29}$$

The nth relation will be obtained in a symmetrical form with respect to the coefficients a_{1i}, a_{2ij}. Let us introduce the sums of the coefficients

$$S_1 = \sum_{i=1}^{n} a_{1i}, \tag{30}$$

$$S_2 = \sum_{i=1}^{n} \sum_{j,j>i} a_{2ij}. \tag{31}$$

Summing the numerators and the denominators of the fractions (29), we have

$$f(\mathrm{H}) = \frac{\sum_{i=1}^{n}(1 - a_{1i})}{\sum_{i=1}^{n} \sum_{j,j\neq i} a_{2ij}} = \frac{n - S_1}{2S_2}. \tag{32}$$

Eliminating $f(\mathrm{H})$ between equations (26) and (32), we obtain

$$4a_0 S_2 = (S_1 - n)(S_1 + n - 2). \tag{33}$$

The number of coefficients and the number of relations indicated above are maximal values that may be reduced for symmetry reasons. In section 4, the relations (29) and (33) will be examplified in several common cases.

(b)   There are no other relations between the coefficients a_0, a_{1i}, a_{2ij} than the preceding ones (or their algebraic combinations). Indeed, let us choose arbitrary substituent parametrizations $g_i(\mathrm{X})$, each of them being equal to zero for $\mathrm{X} = \mathrm{H}$, and arbitrary coefficients b_{2ij}. Then, the property g defined by equation (25) is compatible with the definition (22) of the g_i's. Next, choosing an arbitrary constant $f(\mathrm{H})$, we can define with equation (20) a property f which satisfies equation (5), and whose coefficients a_0, a_{1i}, a_{2ij} are submitted to no other conditions than equations (26), (27).

Let us incidentally notice that, except their common value (19) for $X = H$, there are no general relationships between the parametrizations f_i (since the g_i's can be chosen arbitrarily). In particular, two parametrizations f_i and f_j are not necessarily linked by an affine relationship.

(c)   For a linear model without cross-term, we simply set $a_{2ij} = 0$ in equations (26), (27), which yields $a_{1i} = 1$ for any i, and $a_0 = (1 - n) f(H)$.

4.   Two or three substituents in equivalent or nonequivalent positions

(a) We first consider two positions of substitution, which can be equivalent or non-equivalent, with respect to the measured property f. For two substituents, there is only one coefficient $a_{2ij} = a_2$. The two coefficients a_{1i} are equal, since equation (27) gives

$$a_{11} = 1 - f(H)a_2 = a_{12}. \tag{34}$$

Equation (33) simplifies into

$$a_0 a_2 = a_1(a_1 - 1). \tag{35}$$

(b)   For three substituents there are seven coefficients, $a_0, a_{11}, a_{12}, a_{13}, a_{2,12}, a_{2,23}, a_{2,31}$, linked by the three relations

$$\frac{1 - a_{11}}{a_{2,12} + a_{2,31}} = \frac{1 - a_{12}}{a_{2,23} + a_{2,12}} = \frac{1 - a_{13}}{a_{2,31} + a_{2,23}} \tag{36}$$

and

$$4a_0 S_2 = (S_1 - 3)(S_1 + 1). \tag{37}$$

When the positions of substitution are equivalent, equation (5) can be written as

$$\begin{aligned} f(X, Y, Z) = a_0 &+ a_1\big[f(X) + f(Y) + f(Z)\big] \\ &+ a_2\big[f(X)f(Y) + f(Y)f(Z) + f(Z)f(X)\big], \end{aligned} \tag{38}$$

and equation (37) simplifies into

$$4a_0 a_2 = (a_1 - 1)(3a_1 + 1). \tag{39}$$

Since equation (38) can also be used to treat the case of two substituents only (by setting $Z = H$), we should verify that equations (35) and (39) are compatible. Indeed, when we fix $Z = H$ in equation (38), we get an equation for two substituents only, whose coefficients a_0', a_1', a_2',

$$a_0' = a_0 + a_1 f(H) + a_2 f^2(H), \tag{40}$$
$$a_1' = a_1 + a_2 f(H), \tag{41}$$
$$a_2' = a_2, \tag{42}$$

differ from those a_0, a_1, a_2 of equation (38), and satisfy equation (35).

5. Examples chosen in acid–base equilibria

In this section, and the following one, we shall apply the previous results to experimental values of several physicochemical properties. Our purpose is not to study these physicochemical properties in detail, but only to verify the validity of equations (29), (33) (or their particular cases (36), (37), (39)).

(i) The dissociation constant of carboxylic acids is an example of molecular property f depending on the substituents. Let us consider a family of mono- and disubstituted acetic acids $CH_2X–COOH$ or $CHXY–COOH$, whose pK values (denoted by pK_X or pK_{XY}) have been measured (table 1). If we choose the monosubstituted value $\lambda_X = pK_X$ as a parameter describing the substituent X, we have equation (5):

$$pK_{XY} = a_0 + a_1(pK_X + pK_Y) + a_2 pK_X pK_Y; \tag{43}$$

pK_X and pK_Y are multiplied by the same coefficient a_1, as shown in section 4 (the equivalence of the geminal positions of X and Y, with respect to the measured property pK, need not be put forward here).

Using all the mono- and disubstituted values given in table 1, the coefficients a_0, a_1, a_2 can be optimized in equation (43). When the values obtained, $a_0 = -0.9784$, $a_1 = 0.1919$, $a_2 = 0.1731$, are substituted into equation (43), we obtain a satisfactory model, since the standard deviation on f is 0.11. These coefficients approximately satisfy equation (35), since they give $a_0 a_2 = -0.1694$ and $a_1(a_1 - 1) = -0.1550$ (ratio 1.09). Since equation (35) expresses the compatibility of equations (26) and (27), its best testing consists in comparing the values of $f(H)$ drawn independently from equations (26) and (27). These values, 4.528 and 4.668, are close, and close from the pK of acetic acid (4.756 [12,13]) which is a measure of $f(H)$.

(ii) Similarly, the dissociation constants of the trisubstituted acids can be modelled by an equation of the form (44):

$$pK_{XYZ} = a_0 + a_1(pK_X + pK_Y + pK_Z)$$
$$+ a_2(pK_X pK_Y + pK_Y pK_Z + pK_Z pK_X). \tag{44}$$

By linear regression, we obtain the coefficients $a_0 = 0.5664$, $a_1 = -0.4695$, $a_2 = 0.1619$. Equation (39) is only very approximately satisfied, since we have $4a_0 a_2 = 0.3668$ and $(a_1 - 1)(3a_1 + 1) = 0.6004$ (ratio 1.63). This result is explained by the bad quality of the model (the standard deviation on f is 0.87, due namely to the abnormal molecules (CH_3, CN, CN) and (Ph, OH, F)). However, the compatibility of the equations (26) and (27) is good enough, since the values of $f(H)$ drawn independently from these equations, 4.383 and 4.538, are close (and close to 4.756).

Table 1

pK of mono-, di-, and trisubstituted acetic acids CXYZ–COOH in water, at 25°C (except otherwise stated). The experimental values pK^{exp} of the mono- and disubstituted acids are used to determine a_0, a_1, a_2 by regression in equation (43); those of the mono- and trisubstituted acids are used to determine a_0, a_1, a_2 in equation (44). Abbreviations: Et = ethyl, iPr = isopropyl, Ph = phenyl.

Monosubstituted acids (Y = H, Z = H)					
X	pK^{exp}	Ref.	X	pK^{exp}	Ref.
CH_3	4.86	[11]	COOH	2.83	[11]
Et	4.83	[11]	CN	2.45	[11]
iPr	4.80	[11]	OH	3.83	[11,12]
Ph	4.28	[11]	NH_2	2.35	[11]
CH_2Ph	4.37	[11]	F	2.59	[12]
CH_2OH	4.51	[11]	Cl	2.85	[11]
CH_2COOH	4.16	[11]	Br	2.90	[12]

Disubstituted acids (Z = H)									
X	Y	pK^{exp}	pK^{calc}	Ref.	X	Y	pK^{exp}	pK^{calc}	Ref.
CH_3	CH_3	4.88	4.98	[11]	iPr	COOH	2.94	2.84	[15]
CH_3	Ph	4.64	4.38	[11]	iPr	NH_2	2.29	2.35	[16]
CH_3	COOH	3.07	2.88	[11]	Ph	Ph	3.94	3.84	[11]
CH_3	CN	2.37	2.48	[13]	CH_2Ph	NH_2	2.16	2.09	[16]
CH_3	OH	3.86	3.91	[14]	CH_2OH	OH	3.52	3.61	[11,16]
CH_3	NH_2	2.34	2.38	[11]	CH_2OH	NH_2	2.19	2.17	[11]
CH_3	Cl	2.83	2.90	[11]	CH_2COOH	OH	3.40	3.31	[14]
Et	Et	4.74	4.91	[15]	CH_2COOH	NH_2	1.99	1.96	[16]
Et	Ph	4.37	4.35	[14]	F	F	1.31	1.18	[15]
Et	NH_2	2.29	2.36	[11,16]	Cl	Cl	1.48	1.52	[11,17]
Et	Cl	2.86	2.88	[11]	Br	Br	1.40	1.59	[15]

Trisubstituted acids					
X	Y	Z	pK^{exp}	pK^{calc}	Ref.
CH_3	CH_3	CH_3	5.03	5.19	[11,12]
Ph	Ph	Ph	3.96	3.44	[13]
CH_3	CH_3	CN	2.422	2.53	[13]
CH_3	Ph	CN	2.290	2.12	[13]
CH_3	CN	CN	−2.8	0.81	[13]
CH_3	CH_3	CH_2COOH	3.77	4.42	[13] (meso)
CH_3	CH_3	CH_2COOH	3.93	4.42	[13] (racemic)
CH_2Ph	Ph	CH_2COOH	3.69	3.41	[13]
Ph	Ph	CH_2COOH	3.48	3.33	[13] (meso)
Ph	Ph	CH_2COOH	3.58	3.33	[13] (racemic)
CH_2OH	CH_2OH	CH_2OH	4.460	4.09	[13]

Table 1
(Continued)

Trisubstituted acids

X	Y	Z	pK^{exp}	pK^{calc}	Ref.
CH_3	CH_3	COOH	3.17	2.95	[13]
CH_3	Et	COOH	2.86	2.93	[13]
Et	Et	COOH	2.151	2.91	[13]
iPr	iPr	COOH	2.124	2.86	[13]
CH_3	CH_3	OH	4.04	4.06	[13] (18°C)
CH_3	Et	OH	3.991	4.03	[13] (18°C)
Et	Et	OH	3.804	4.00	[13] (18°C)
CH_3	Ph	OH	3.467	3.51	[18]
Ph	Ph	OH	3.036	3.02	[18]
CH_3	CH_2COOH	OH	3.5	3.40	[16]
CH_2COOH	CH_2COOH	OH	3.128	2.82	[13,18]
CH_3	CH_3	NH_2	2.36	2.42	[11]
iPr	NH_2	OH	2.55	1.67	[13]
CH_2COOH	OH	NH_2	1.91	1.33	[13]
Ph	OH	F	4.244	1.60	[13]
F	F	F	0.50	0.18	[13]
F	F	Cl	0.46	0.27	[13]
CH_3	CH_3	Cl	2.975	2.97	[13]
CH_3	Cl	Cl	2.06	1.41	[13]
Cl	Cl	Cl	0.52	0.50	[13]
CH_3	Br	Br	1.48	1.49	[13]
Br	Br	Br	−0.147	0.57	[13]

6.   An example in NMR

The NMR vicinal coupling constant J between two hydrogen atoms is another example of property f which has been studied as a function of the substituents. We consider a family of ethanes CH_2X–CH_3 or $CHXY$–CH_3 bearing one substituent X or two geminal substituents X, Y. The corresponding $H \cdots H$ coupling constant (denoted by J_X or J_{XY}), taken from the literature, or newly measured, are given in table 2. Choosing the parameter $\lambda_X = J_X$ to represent the substituent X, we have

$$J_{XY} = a_0 + a_1(J_X + J_Y) + a_2 J_X J_Y. \tag{45}$$

By regression in equation (45), using all the measured coupling constants of table 2, we obtain $a_0 = -178.69$, $a_1 = 23.621$, $a_2 = -2.9972$. These coefficients satisfy equation (35), since they give $a_0 a_2 = 535.57$ and $a_1(a_1 - 1) = 534.33$ (ratio 1.002). The compatibility of equations (26) and (27) is excellent since the values of $f(H)$ drawn from them are 7.556 and 7.547, respectively. These results are good, when compared to the mean quality of the regression (the standard deviation on J is 0.23, due namely to the abnormal molecules (OAc, C(=O)Me) and (F, F)).

Table 2

Vicinal $H \cdots H$ NMR coupling constants in mono- or disubstituted ethanes $CHXY–CH_3$. Solvent $CDCl_3$. The coupling constants referenced with *a* were measured by J.-C. Ziegler on a 400 MHz spectrometer at 25°C temperature, and at a 2% concentration in weight (solid) or in volume (liquid). The experimental values J^{exp} of the mono- and disubstituted ethanes are used to determine a_0, a_1, a_2 by regression in equation (43). Abbreviations: Me = methyl, Et = ethyl, iPr = isopropyl, Ac = acetyl, Bz = benzyl, Ts = tosyl.

Monosubstituted ethanes					
X	J_X^{exp}	Ref.	X	J_X^{exp}	Ref.
Li	8.4	[19]	OH	7.025	[22]
CH_3	7.35	[20–22]	OEt	7.02	[20–22]
Et	7.42	[20,22]	OAc	7.145	*a*
Ph	7.616	*a*	OBz	7.136	*a*
CN	7.63	[20–22]	OTs	7.126	*a*
CHO	7.39	[20,22]	ONO_2	7.06	[23,24]
C(=O)Me	7.332	*a*	SH	7.34	[20–22]
C(=O)Et	7.38	[20–22]	F	7.00	[20,22]
COOH	7.56	[20–22]	Cl	7.26	[20–22]
COOEt	7.58	[20,22]	Br	7.35	[19,20,22]
NH_2	7.12	[20–22]	I	7.47	[19,20,22]
NO_2	7.37	[20–22]			

Disubstituted ethanes				
X	Y	J_{XY}^{exp}	J_{XY}^{calc}	Ref.
CH_3	CH_3	6.61	6.62	[20,22]
CH_3	Et	6.64	6.73	[20,22]
CH_3	Ph	6.92	7.03	[20,22]
CH_3	CN	7.05	7.07	[20–22]
CH_3	CHO	7.07	6.68	[20,22]
CH_3	C(=O)iPr	6.88	6.67	[20–22]
CH_3	COOH	7.00	6.95	[20–22]
CH_3	NH_2	6.26	6.25	[21,22]
CH_3	NO_2	6.65	6.65	[20–22]
CH_3	OH	6.13	6.16	[20,22]
CH_3	OiPr	6.11	6.09	[22]
CH_3	ONO_2	6.20	6.16	[23,24]
CH_3	OAc	6.266	6.29	*a*
CH_3	OBz	6.254	6.28	*a*
CH_3	OTs	6.250	6.26	*a*
CH_3	SH	6.68	6.60	[20–22]
CH_3	F	6.11	6.06	[20,22]
CH_3	Cl	6.49	6.48	[20,22]
CH_3	Br	6.59	6.62	[20,22]
CH_3	I	6.77	6.81	[20,22]

Table 2
(Continued)

Disubstituted ethanes				
X	Y	J_{XY}^{exp}	J_{XY}^{calc}	Ref.
Et	Li	7.88	8.18	[22]
Et	Cl	6.54	6.61	[20,22]
Et	Br	6.67	6.73	[20,22]
Et	I	6.82	6.90	[20,22]
Ph	CHO	7.07	7.07	[20,22]
Ph	COOH	7.174	7.20	*a*
Ph	NH_2	6.621	6.86	*a*
Ph	OH	6.450	6.71	*a*
Ph	OAc	6.607	6.88	*a*
Ph	OBz	6.566	6.87	*a*
Ph	OTs	6.404	6.87	*a*
Ph	Cl	6.832	6.97	*a*
CN	Cl	7.10	6.94	[23,24]
CN	Br	7.12	7.07	[20,22]
C(=O)Me	Cl	6.900	6.44	*a*
COOH	COOH	7.373	7.14	*a*
COOH	SH	7.10	7.00	[23,24]
COOH	Cl	6.99	6.87	[20,22]
COOH	Br	6.95	6.95	[20,22]
COOEt	C(=O)R	7.161	6.98	*a*
COOEt	COOEt	7.3	7.16	[20,22]
COOEt	OAc	7.007	6.76	*a*
COOEt	OBz	7.054	6.75	*a*
COOEt	OTs	6.931	6.74	*a*
COOEt	Br	6.939	6.97	*a*
OEt	OEt	5.35	5.24	[20,22]
OAc	C(=O)Me	7.086	6.25	*a*
F	F	4.52	5.14	[20,22]
Cl	Cl	6.04	6.31	[20,22]
Cl	I	6.5	6.70	[23,24]
Br	Br	6.35	6.62	[23,24]
I	I	7.0	6.96	[23,24]

7.   Conclusion

All the alternative sets (λ_i) of substituent parametrizations which are compatible with a linear plus cross-term model (4) are affine transforms (6) of one another. For arbitrary parametrizations λ_i there is no general relationship between the coefficients a_0, a_{1i}, a_{2ij} of the model. In the case of equivalent positions of substitution, the expression (18) is invariant in a change of parametrization.

The self-parametrization (f_i) (the case when the monosubstituted value $f_i(X)$ of f is chosen as the parameter representing the substituent X in position i) is one of the

alternative sets (λ_i). In this case, we have shown that the coefficients a_0, a_{1i}, a_{2ij} of equation (5) are linked by the n important relations (29) and (33).

For two substituents in equivalent positions, the coefficients a_0, a_1, a_2 are connected by the simple relation (35). It is a nonintuitive result that, for two nonequivalent positions, the coefficients a_{1i} and a_{1j} still coincide, and that the relation (35) is still valid. For three substituents in equivalent positions, the coefficients a_0, a_1, a_2 are linked by equation (39). We have verified that these relations are well or roughly satisfied by the coefficients a_0, a_1, a_2 optimized from measured acid–base dissociation constants, or from NMR coupling constants.

Having a set of measured values of a molecular property f, one has two possibilities for creating a linear plus cross-term model: either (i) take the relations (35) or (39) into account to reduce the number of coefficients to be optimized (this is interesting when one has few experimental points); or (ii) ignore these relations and optimize all the coefficients as if they were independent, to get a better model (more coefficients give a more accurate model). However, even when one makes the choice (ii), it is essential to be aware that the coefficients are theoretically dependent, and to know the relations linking them.

In this article, the study of self-parametrization was restricted to the linear plus cross-term models. It is clear that similar reasonings can be applied to other forms of model equations, showing also that the coefficients are interdependent, but with different relations.

We have chosen the substituent H as the reference substituent in the definition of the f_i's, and throughout this article. However, any type of substituent can be chosen instead of H, with the same results. Moreover, the reasoning, and the equations obtained, are still valid when the notion of substituent is extended into any structural element which modifies the environment of the active site of the molecule. For example, a chain of atoms bonded on its two ends to the substrate can be considered as a substituent.

We have assumed in the introduction that f was a scalar property. Self-parametrization can also be considered for a vector property $f = (f^{(1)}, \ldots, f^{(p)})$. Then, several parameters $\lambda_i^{(k)}(X) = f_i^{(k)}(X)$ will represent the influence of a substituent X in position i, and tensor coefficients a will be needed.

References

[1] N.B. Chapman and J. Shorter, eds., *Advances in Linear Free Energy Relationships* (Plenum Press, New York, 1972).

[2] N.B. Chapman and J. Shorter, eds., *Correlation Analysis in Chemistry. Recent Advances* (Plenum Press, New York, 1978).

[3] M. Karelson, V.S. Lobanov and A.R. Katritzky, Chem. Rev. 96 (1996) 1027–1043.

[4] C. Hansch, D. Hoekman and H. Gao, Chem. Rev. 96 (1996) 1045–1075.

[5] M. Charton and B.I. Charton, eds., *Advances in Quantum Structure–Property Relationships*, Vol. 2 (JAI Press, Stamford, CT, 1999).

[6] J.A. MacPhee, A. Panaye and J.-E. Dubois, Tetrahedron 34 (1978) 3553–3562.

[7] C. Hansch, in [2], pp. 397–438.

[8] O. Exner, in [2], pp. 439–540.

[9] S. Pietri, M. Miollan, S. Martel, F. Le Moigne, B. Blaive and M. Culcasi, J. Biol. Chem. 275 (2000) 19505–19512.

[10] I.V. Shpanko, Org. React. 31 (1997) 177–180.

[11] D.R. Lide, ed., *Handbook of Chemistry and Physics*, 78th ed. (CRC Press, Boca Raton, 1997).

[12] A. Albert and E.P. Serjeant, *The Determination of Ionization Constants. A Laboratory Manual*, 3rd ed. (Chapman and Hall, London, 1984).

[13] J.A. Dean, *Handbook of Organic Chemistry* (McGraw-Hill, New York, 1987).

[14] R.C. Weast, ed., *Handbook of Chemistry and Physics* (CRC Press, Boca Raton, 1992).

[15] S. Patai, *The Chemistry of Carboxylic Acids and Esters* (Interscience, London, 1969).

[16] R.M.C. Dawson, D.C. Elliot, W.H. Elliot and K.M. Jones, *Data for Biochemical Research* (Oxford University Press, Oxford, 1969).

[17] C. Moreau, Bull. Soc. Chim. Fr. (1968) 31.

[18] S. Budavari, ed., *The Merck Index, an Encyclopedia of Chemicals, Drugs, and Biologicals* (Merck and Co., Whitehouse Station, NJ, 1996).

[19] W.J. Colucci, S.J. Jungk and R.D. Gandour, Magn. Res. Chem. 23 (1985) 335–343.

[20] L.A. Donders, Computation and analysis of NMR spin–spin interactions, Thesis, Rijkuniversiteit, Leiden (1989).

[21] C. Altona, J.H. Ippel, A.J.A. Westra Hoekzema, C. Erkelens, M. Groesbeek and L.A. Donders, Magn. Reson. Chem. 27 (1989) 564–576.

[22] C. Altona, Magn. Res. Chem. 32 (1994) 670–678.

[23] O. Yamamoto, K. Someno, N. Wasada, J. Hiraishi, K. Hayamizu, K. Tanabe, T. Tamura and M. Yanagisawa, Analyt. Sci. 4 (1988) 233–239.

[24] O. Yamamoto, K. Someno, N. Wasada, J. Hiraishi, K. Hayamizu, K. Tanabe, T. Tamura and M. Yanagisawa, Spectral data base system, Internet, Japan (1998).