

Range of validity of the Hammett equation: acidity of substituted ethynylbenzenes†

Stanislav Böhm,^a Patrik Pařík^b and Otto Exner^{*c}

Received (in Durham, UK) 7th September 2005, Accepted 16th December 2005

First published as an Advance Article on the web 20th January 2006

DOI: 10.1039/b512698c

Acidities of 19 *meta*- and *para*-substituted ethynylbenzenes were calculated at the B3LYP/6-311+G(d,p) level and correlated within the framework of the Hammett equation with the calculated acidities of equally substituted benzoic acids. The substituent effects were decomposed in terms of isodesmic reactions into those operating in the anions and in the uncharged molecules. Characteristic deviations from the Hammett equation were found for *para*-substituents, both for acceptors and donors; the former can be interpreted by the resonance formula only with an electron sextet. With reference to the series of ionization reactions investigated previously, it was possible to reinvestigate the validity of the Hammett equation on the basis of calculated reaction energies using a more homogeneous data set than had been ever accessible from the experimental reactivities. The equation was fulfilled for all *meta*-substituents with a higher accuracy than commonly attainable with the experimental data. When *para*-substituents were included, deviations occurred according to the character of the functional group: When this group was an acceptor, the donor substituents showed deviations and *vice versa*. Another series of reactions proceeding between uncharged groups bonded directly on the benzene ring was investigated in the same way: The Hammett equation held with a similar precision, although its original range of validity was surpassed. The properties of a set of common substituents were investigated by principal component analysis and cluster analysis. There is a fundamental difference between uniform acceptors and more discriminated donors but clustering is not so strong to depreciate common statistical analysis.

The Hammett equation has become popular, in spite of its restricted applicability, as the first attempt to predict reactivity by means of an empirical formula. For many years, it has remained the most general and simplest structure–property relationship.^{1,2} In energy terms, it is expressed by eqn (1).

$$E(X) - E(H) = \rho\sigma_{m,p} \quad (1)$$

The symbol $E(X)$ may mean the reaction energy or enthalpy but in most cases it stands for the Gibbs energy, or activation Gibbs energy, in a series of reactions of *meta*- or *para*-substituted benzene derivatives. The empirical parameter σ_m or σ_p characterizes the substituent X, the parameter ρ is pertinent to the given reaction, $E(H)$ relates to the unsubstituted compound ($X = H$). In simple terms of organic chemistry, eqn (1) means that the variable substituents X raise changes of reactivity always in the same succession: for instance the substituent effect

of 4-CO₂CH₃ is always twice the effect of 4-Cl. Eqn (1) has the main importance in interpreting the values of the parameters² σ and ρ ; less in predicting the unknown values of E . Nevertheless, it serves still as reference when investigating reactivities or activities of benzene derivatives.³

The exact range of validity is of central importance for any empirical relationship² and numerous attempts were made to delimit it for eqn (1) with more precision.^{2,4} In kinetics, the deviations encountered were mostly attributed to complex reaction mechanism.² A more fundamental limitation was found in the deviations of some *para*-substituents when they were directly conjugated with the functional group.^{2,4} The most correct solution is strict limitation of the validity range to reactions proceeding in the side chain at a non-conjugated reaction center,² or only the conjugated *para*-substituents can be excluded.^{4d} However, a purely formal way out was chosen most often, which defines special constants (denoted σ_p^+ or σ_p^-) appropriate for certain substituents in certain reactions.⁵ These limitations and corrections lead ultimately to the hypothesis that *para*- and *meta*-derivatives should be treated separately in two equations. The hypothesis was not verified on the basis of solution reactivities⁶ but obtained some support later from the gas-phase acidities and basicities.⁷ Many attempts have been made to improve eqn (1) by adding further terms;^{5a,8} they were critically evaluated.^{7,9} Further improvement was searched in dividing the substituents into classes¹⁰ or even modifying the linear form.¹¹ A complete failure of the

^a Institute of Organic Chemistry, Prague Institute of Chemical Technology, CZ-16628 Praha 6, Czech Republic

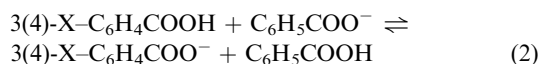
^b Department of Organic Chemistry, Faculty of Chemical Technology, University of Pardubice, CZ-53210 Pardubice, Czech Republic

^c Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, CZ-16610 Praha 6, Czech Republic

† Electronic supplementary information (ESI) available: Tables S1 giving MP2 calculated energies of some substituted phenylacetic acids, S2 giving energies of substituted benzyl alcohols, S3–S6 summarizing calculated reaction energies as source matrices for testing the Hammett equation and Fig. S1 displaying a tree diagram from the cluster analysis of substituents. See DOI: 10.1039/b512698c

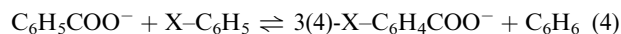
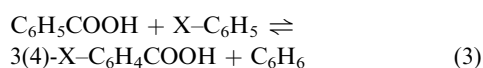
Hammett equation was observed only in rare cases, mostly for reactions of uncharged molecules.^{12,13}

In order to anchor the scale of constants σ , dissociation of substituted benzoic acids in water was chosen as reference reaction.¹ The relative p*K* values are proportional to the Gibbs energies $\Delta_2G^\circ(298)$ of the isodesmic reaction, eqn (2), in which only a proton is transferred.



Available p*K* values are not quite accurate standards¹⁴ and the effects of certain substituents may be spoiled by specific solvent effects. More consistent standards could be obtained when eqn (2) was related to isolated molecules, investigated either experimentally in the gas phase¹⁵ or by quantum chemical calculations.^{16–18}

For theoretical interpretation, eqn (2) has still another shortcoming in that it expresses only the difference of two effects: interaction of the groups X and COO[−] in the anion XC₆H₄COO[−] and of X and COOH in the molecule XC₆H₄COOH. Significant progress was achieved when these effects were evaluated separately as the reaction energies Δ_3E and Δ_4E of the isodesmic reactions,¹⁹ eqn (3) and (4).



All the reactions, eqn (2)–(4), meet also the definition of a homodesmotic reaction²⁰ but eqn (3) and (4) exceed the above-

mentioned range of validity since the reaction proceeds directly on the benzene ring instead of in the side chain. The energies Δ_3E and Δ_4E can be obtained from the experimental enthalpies of formation but with a low accuracy.²¹ More efficiently they were obtained from quantum chemical calculations^{17,18} that are relatively accurate for isodesmic reactions.²² It turned out that the substituent effect is more important in the anions (Δ_4E) than in the acid molecules (Δ_3E).^{17,18} Both cooperate in their influence on the acidity (Δ_2E) but are not exactly proportional: Δ_3E shows some irregularities that are partly transferred into Δ_2E .¹⁷ Acidities of substituted phenols²³ and basicities of several classes of compounds^{24–27} were analyzed similarly.

The subject of this paper is a revision of the Hammett equation on isolated molecules using calculated reaction energies. The study of isolated molecules gives important information, complementary to that obtained experimentally in solution. Not only are some effects of specific solvation eliminated, but also reactions of a new type can be added. A reaction like that in eqn (2) can hardly be imagined in solution. On the other hand, ionization reactions like eqn (1) were studied in solution and in isolated molecules with essentially concordant results.¹⁵

From previous work we had at our disposal the acidities¹⁷ and basicities²⁴ of substituted benzoic acids and the basicities of benzonitriles²⁵ and nitrobenzenes;²⁶ three series were taken from Wiberg^{18,23} and completed. Here we added an important series, the acidities of substituted ethynylbenzenes **1** (Table 1). For the reactions of uncharged molecules, similar to eqn (3), we had at our disposal three series from our work^{17,25,26} and

Table 1 DFT energies of substituted ethynylbenzenes **1a–1y** and energies of the isodesmic reactions

	Substituent	$\Delta E(\text{DFT})$ a. u.	$\Delta E(\text{DFT})$ anion a. u.	Isodesmic energies (kJ mol ^{−1})		
				Δ_5E	Δ_6E	Δ_7E
1a	H	−308.4772460	−307.8775790	0.00	0.00	0.00
1b	3-CH ₃	−347.8049265	−347.2041402	2.94	−0.37	2.57
1c	3-CH ₂ Cl	−807.4261280	−806.8331785	−17.64	1.63	−16.01
1d	3-CF ₃	−645.6254389	−645.0381023	−32.37	3.65	−28.72
1e	3-CHO	−421.8338961 ^a	−421.2465534 ^a	−32.36	3.10	−29.25
1f	3-CN	−400.7420360	−400.1585675	−42.53	4.55	−37.98
1g	3-NH ₂	−363.8538659	−363.2518600	6.14	−0.60	5.54
1h	3-N(CH ₃) ₂	−442.4804924	−441.8772752	9.32	−1.14	8.18
1i	3-NO ₂	−513.0386209	−512.4566979	−46.59	5.24	−41.35
1j	3-OH	−383.7243908 ^b	−383.1275523 ^b	−7.43	0.65	−6.78
1k	3-F	−407.7452418	−407.1525676	−18.36	2.31	−16.05
1l	3-Cl	−768.0992471	−767.5081198	−22.42	2.47	−19.95
1m	3-SO ₂ CH ₃	−896.4448753 ^c	−895.8613108 ^c	−42.28	4.03	−38.25
1n	4-CH ₃	−347.8051573	−347.2033020	5.75	−0.97	4.77
1o	4-CH ₂ Cl	−807.4265337	−806.8364219	−25.09	0.56	−24.52
1p	4-CF ₃	−645.6256475	−645.0413296	−40.30	3.10	−37.20
1q	4-CHO	−421.8345788	−421.2537605	−49.49	1.31	−48.18
1r	4-CN	−400.7426885	−400.1636780	−54.23	2.84	−51.40
1s	4-NH ₂	−363.8550843	−363.2480338	19.39	−3.80	15.58
1t	4-N(CH ₃) ₂	−442.4820335	−441.8751015	19.07	−5.19	13.89
1u	4-NO ₂	−513.0395093	−512.4667920	−70.76	2.90	−67.85
1v	4-OH	−383.7251172	−383.1226299	7.40	−1.26	6.15
1w	4-F	−407.7455862	−407.1498510	−10.32	1.40	−8.92
1x	4-Cl	−768.0996169	−767.5072228	−19.09	1.50	−17.60
1y	4-SO ₂ CH ₃	−896.4451666 ^c	−895.8656246 ^c	−52.84	3.27	−49.57

^a The two conformations of the C1–C(O) bond are populated *ap*:*sp* 59:41 in the neutral molecule and 95:5 in the protonated form. ^b The two conformations of the C1–C(O) bond are populated *ap*:*sp* 41:59 in the neutral molecule and 6:94 in the protonated form. ^c Conformation of the C1–S bond with the CH₃ group above the ring plane.

four from the literature,^{16d} in addition to ethynylbenzenes. Altogether we had two complete sets, each with 8 reaction series and with the same 19 substituents; such a complete set has never been available from experimental data.^{4,7,8} The main problems are firstly the validity of the equation separately for *meta*- and *para*-substituents and secondly the separation of substituent effects in the neutral molecules and in ions, like in the reactions of eqn (2) and (3).

Calculations

Energies of substituted ethynylbenzenes **1a–1y** (Table 1) and of their deprotonated forms were calculated by density functional theory (DFT)²⁸ at the B3LYP/6-311+G(d,p) level using the GAUSSIAN 03 program.²⁹ Corrections for the zero-point energy were not introduced. Planarity or any symmetry was never anticipated. All reasonable conformations were taken into consideration and calculations were started always from the pertinent near structures; this concerns particularly the two possible orientations of some substituents in *meta*-derivatives. All structures were checked by vibrational analysis and behaved as energy minima. The energies $E(\text{DFT})$ are listed in Table 1 and some geometrical parameters of selected compounds are shown in Table 2. The reaction energies Δ_5E to Δ_7E (Table 1) related to eqn (5)–(7) were obtained from these values and from $E(\text{DFT})$ of the compounds $\text{C}_6\text{H}_5\text{X}$ calculated previously with the same theoretical model.²¹

In the cases where several conformers coexist, the values of Δ_5E to Δ_7E were corrected to relate to the equilibrium mixture at 298 K. The equilibrium constant and population of conformers were estimated with the assumption $\Delta G^\circ(298) \cong \Delta E(\text{DFT})$; from the population the effective energies of the equilibrium mixture were calculated. The values of Δ_5E to Δ_7E in Table 1 have been calculated in this way but the differences compared to the lowest-energy conformer are minute.

Energies of some substituted phenylacetic acids **2a** and **2b** and of substituted benzyl alcohols **3a** to **3e** were calculated, closely following the conditions used previously^{18,23} for the other derivatives, *viz.* at the levels MP2/6-311++G(d,p)//MP2/6-311+G(d) or MP2/6-311++G(2df,2pd)//MP2/6-311+G(d), respectively. The conformations of **3a–e** were investigated in more detail than previously.²³ The results are given in Tables S1 and S2 (ESI†).

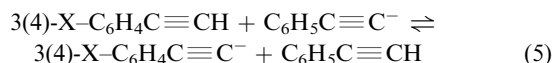
Table 2 Calculated bond lengths in selected substituted ethynylbenzenes (in Å)

Compound	Substituent	Uncharged molecule		Anion	
		C≡C	C(1)–C	C≡C	C(1)–C
1t	4-N(CH ₃) ₂	1.206	1.425	1.247	1.417
1s	4-NH ₂	1.206	1.426	1.247	1.419
1v	4-OH	1.205	1.427	1.247	1.418
1a	H	1.205	1.428	1.248	1.415
1p	4-CF ₃	1.204	1.427	1.258	1.405
1q	4-CHO	1.205	1.426	1.251	1.402
1r	4-CN	1.205	1.426	1.250	1.404
1u	4-NO ₂	1.204	1.426	1.252	1.399

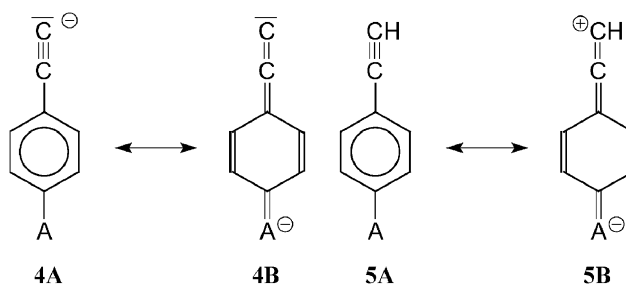
Results and discussion

Acidity of substituted ethynylbenzenes

Substituted ethynylbenzenes **1a–1y** (Table 1) were chosen as a model series, in which characteristic deviations from the Hammett equations were anticipated. The relative acidity is defined as the Gibbs energy Δ_5G° of the reaction of eqn (5).



We calculated the pertinent DFT energies Δ_5E (Table 1, column 5), since it had been proven several times that calculated enthalpies or Gibbs energies do not give better agreement with the experimental acidities.^{17,30} If the Hammett equation were valid exactly, Δ_5E would be proportional to the acidities of the benzoic acids¹⁷ Δ_2E . Fig. 1 reveals that it holds for all *meta*-derivatives and for *para*-derivatives with donor substituents. According to the correlation coefficient R , the correlation is closer than commonly attainable with reactivities in solution. Deviations of *para*-acceptors are very significant. The explanation by through-conjugation is suggested: resonance in the anion **4A** ↔ **4B** should be stronger than in the uncharged molecule **5A** ↔ **5B**. This is in accord with the common rules of resonance (separation of charges in **5B**).



The structures **4B** and **5B** are atypical having an electron sextet on carbon, and their representation should be minute. Nevertheless, their significance is supported by the calculated bond lengths. The single bond C(1)–C_{sp} in the anion is shortened in the presence of an acceptor (Table 2, last column) as required by the structure **4B**. Lengthening of the triple bond (last but one column) is much less significant and need not be observable in the experimental values. In the uncharged molecules the variability of bond lengths is much smaller, almost negligible. It was observed several times that the resonance formulae predict the geometry well in the essential features but not in all details: in particular lengthening of the multiple bonds is less evident than shortening of the single bonds.^{31,32} There is also no exact parallelism between the resonance effects on the energy and on the geometry.^{26,32,33}

Fig. 1 reveals very clearly the limited validity of the Hammett equation. Note however that the deviations could escape attention if the data were treated in the usual way referring only to simple statistics. The value of the correlation coefficient (Fig. 1, the figures in square brackets) would be empirically classified^{4a} as ‘good fit’, almost ‘excellent fit’. The example shows nicely how misleading the simple statistical treatment may be. More sophisticated statistics would reveal

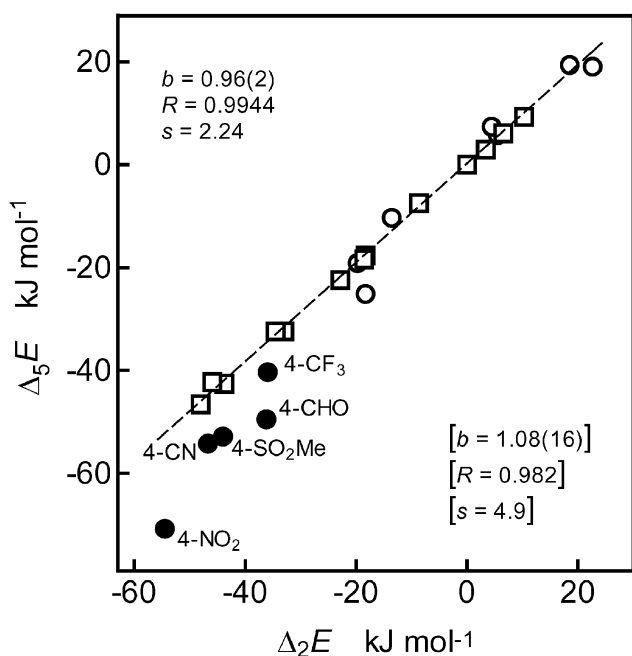


Fig. 1 Hammett plot of calculated acidities of *meta*- and *para*-substituted ethynylbenzenes Δ_5E vs. the acidities of equally substituted benzoic acids Δ_2E : \square *meta*-substituents, \circ *para* donor substituents, \bullet *para* acceptor substituents; the regression line and the statistics relate to groups $\square + \circ$, the figures in square brackets relate to all points.

the difference between *meta*- and *para*-substituents but it is evident from the graph on first sight.

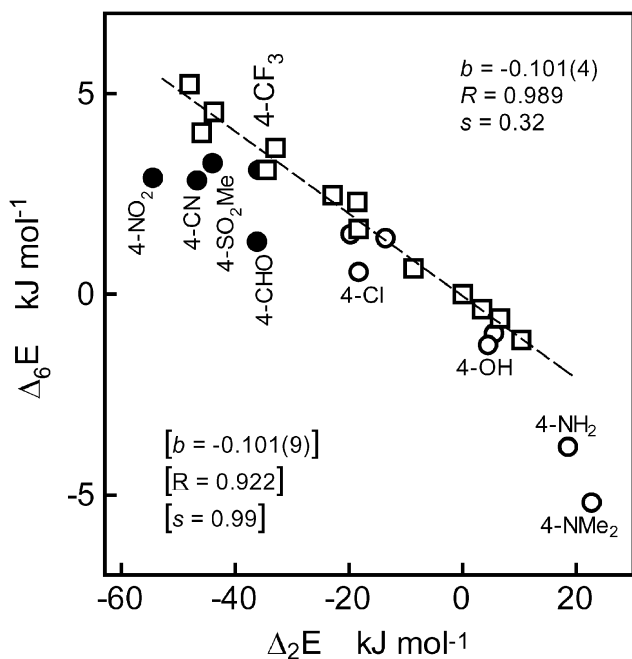
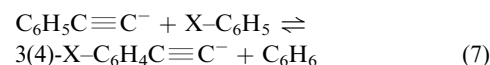
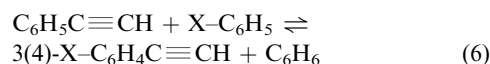
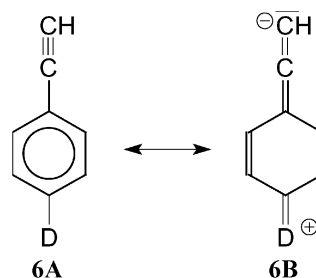


Fig. 2 Hammett plot of calculated energies Δ_6E of the isodesmic reaction eqn (6) for substituted ethynylbenzenes vs. the relative acidities Δ_2E of equally substituted benzoic acids; definition of points as in Fig. 1; the regression line and the statistics relate only to *meta*-substituents, the figures in square brackets relate to all points.

Estimation of the substituent effects separately in the uncharged ethynylbenzenes and in their anions is based on eqn (6) and (7).



The energies Δ_6E extend over a range of only 10 kJ mol^{-1} ; hence their correlations could be less precise. Nevertheless, the dependence on Δ_2E is very good for *meta*-substituents (Fig. 2) while *para*- deviate strongly, both acceptors and donors to the same side. While the deviation of acceptors is explained by the formulae **5A** \leftrightarrow **5B** with a sextet on C, that of donors, **6A** \leftrightarrow **6B**, does not require a sextet and should be relatively strong. Shortening of the C(1)–C_{sp} bond (Table 2, column 4) is very small; nevertheless this bond is longest in unsubstituted **1a**. Viewed from the side of the ethynyl group, this group behaves as a donor towards acceptors (**5B**) and as an acceptor towards donors (**6B**). Similar behaviour was well documented in the case of the ethenyl group.¹³



However strong are the deviations of *para*-substituents in Fig. 2, they could be overseen even in this case when only simple statistics were used: the figures in square brackets would be classified^{4a} as ‘fair fit’.

Correlation of the energies of the anion Δ_7E (Fig. 3) brought nothing new: it is very similar to that of the acidities Δ_5E (Fig. 1). The acidities of **1** are controlled by the effects in the anions: the energies of the uncharged acids are eight times less sensitive to the substituent effects. Similarly the basicities of benzonitriles are controlled by the effects in the cations, effects in the uncharged bases being much less important.²⁵ This behaviour makes the Hammett-type reactions, controlled by polar effects, a very suitable model for an empirical equation. It is greatly different from the behaviour of *ortho*-derivatives with alkyl substituents³⁴ that is controlled by steric effects. These effects are of the same sign in the cations, anions and uncharged species; the resulting effect on acidobasic properties is then difficult to predict.³⁴

Multi-parameter equations

The common practice of improving the Hammett equation makes use of more complex equations.⁸ These procedures were several times reviewed and criticized.^{7,9,25} In our opinion, many of them have not been rigorously established and can be misleading; nevertheless, we tried in this section to test at least two of the most popular of such models on the two series of our data, *viz.* on the acidities Δ_5E and energies Δ_6E . Using the dual substituent parameter (DSP) treatment^{8b} with the

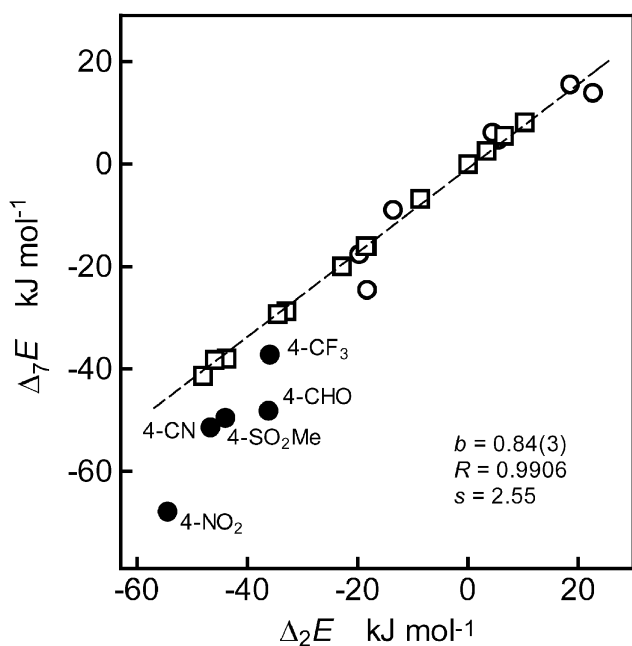


Fig. 3 Hammett plot of calculated energies Δ_7E of the isodesmic reaction eqn (7) for the anions of substituted ethynylbenzenes vs. the relative acidities Δ_2E of equally substituted benzoic acids; definition of points as in Fig. 1; the regression line and the statistics relate only to groups $\square + \circ$.

constants σ_I and σ_R designed for the gas phase,^{15b} we obtained eqn (8) and (9) for the *meta*- and *para*-derivatives, respectively.

$$\Delta_5E(m) = -2.8 \pm 1.7 - (60.2 \pm 4.0)\sigma_I - (33.3 \pm 3.4)\sigma_R$$

$$R = 0.9920 \quad s = 2.72 \quad N = 13 \quad (8)$$

$$\Delta_5E(p) = -6.1 \pm 4.9 - (68.9 \pm 11.6)\sigma_I - (69.8 \pm 9.7)\sigma_R$$

$$R = 0.972 \quad s = 7.80 \quad N = 13 \quad (9)$$

On the first sight the correlation is good; see the correlation coefficients R and standard deviations of the regression coefficients. The negative slopes ρ mean that the acidity is strengthened by electron attracting substituents. However, the fit is greatly different for *meta*- and *para*-derivatives, see particularly the standard deviations s . For *meta*-derivatives the correlation is evidently overparametrized since the fit is not better than in Fig. 1.

Correlation of the energies Δ_6E provided eqn (10) and (11), which exhibit the same features still more clearly; positive slopes ρ mean that the molecule is destabilized by electron attracting substituents. Summarizing, the improvement reached with the DSP model is not proportional to the extended number of parameters.

$$\Delta_6E(m) = 0.14 \pm 0.16 + (6.73 \pm 0.38)\sigma_I + (3.34 \pm 0.31)\sigma_R$$

$$R = 0.9940 \quad s = 0.25 \quad N = 13 \quad (10)$$

$$\Delta_6E(p) = -0.51 \pm 0.54 + (5.13 \pm 1.27)\sigma_I + (6.57 \pm 1.06)\sigma_R$$

$$R = 0.956 \quad s = 0.85 \quad N = 13 \quad (11)$$

The Yukawa–Tsunoo treatment was carried out with standard^{8c} $\sigma_{m,p}$ and with the parameters^{8a} $\Delta\sigma_R^-$ expressing enhanced conjugation of the acceptor substituents. Eqn (12) is more effective than eqn (8) and (9) due to the smaller number

of parameters since *meta*- and *para*-derivatives are treated together. The same holds for eqn (13) when compared with eqn (10) and (11).

$$\Delta_5E = -7.9 \pm 1.4 - (49.4 \pm 3.6)\sigma_{m,p} - (69.8 \pm 15.2)\Delta\sigma_R^-$$

$$R = 0.971 \quad s = 6.20 \quad N = 24 \quad (12)$$

$$\Delta_6E = 0.31 \pm 0.11 + (6.67 \pm 0.29)\sigma_{m,p} - (7.43 \pm 1.23)\Delta\sigma_R^-$$

$$R = 0.982 \quad s = 0.50 \quad N = 24 \quad (13)$$

Somewhat puzzling might be the inverse signs of slopes at $\sigma_{m,p}$ and $\Delta\sigma_R^-$ in eqn (13). The latter term usually expresses that acceptor substituents are conjugated more strongly; here they are conjugated more weakly than donors (Fig. 2). The fundamental difference between *meta*- and *para*-substituents is not reproduced well even by this model.

General test of the Hammett equation

We investigated two sets of data. The set **A** was restricted to acidobasic reactions that formed the most frequent equilibrium data, which have been correlated by the Hammett equation.^{7,8} We used calculated reaction energies of eight series: ionization of substituted benzoic acids,¹⁷ phenylacetic acids¹⁸ [literature data¹⁸ completed as given in Table S1 (ESI[†])] benzyl alcohols [literature data²³ completed as given in Table S2 (ESI[†])], ethynylbenzenes (Table 1) and phenols,²³ and further protonation of benzonitriles,²⁵ benzoic acids²⁴ and nitrobenzenes.²⁶ Note that these calculated values had always been checked by comparison with the available experimental gas-phase acidities. The set **B** was composed from isodesmic reactions of uncharged compounds, similar to eqn (3) or (6). Such reactions do not belong to the range of validity of the Hammett equation as originally defined since the reaction proceeds directly on the benzene ring instead of in the side-chain. Nevertheless such reactions were correlated with modified constants σ with only slightly lowered accuracy.^{2,4} Our set **B** contained eight series always with a constant common group: COOH,¹⁷ C \equiv CH (Table 1), OH,²³ SH,^{16d} NH₂,^{16d} N(CH₃)₂,^{16d} CN,²⁵ and NO₂.²⁶ The set *meta*-OH had to be eliminated due to a very low measure of joint variability. (The substituent effects are too small and hidden in the noise.) The substituents were the same in the sets **A** and **B**: H, CH₃, CF₃, CHO, CN, NH₂, NO₂, OH, F, Cl; each in both the *meta*- and *para*-positions. They belong to the standard substituents most used in practice. For convenience we reproduce our input data as the source matrices in Tables S3 to S6 (ESI[†]). Both contain 8 × 19 items; **B** is filled only to 94% after eliminating the *meta*-OH set. We hoped that we could obtain more significant results than in the previous investigation⁷ based on the experimental gas-phase acidities (matrix 21 × 11 filled to 83%) or in the pioneer study^{4d} based on solution reactivities (matrix 62 × 63 filled to 25%).

Our main problem was whether the *meta*-derivatives fulfil the equation with a significantly better accuracy than the *para*-, either in all or in selected series. We used two main methods: simple linear regression (LR) and principal component analysis (PCA).

LR was based on the conventional choice of the standard reaction series: ionization of benzoic acids was used as the explanatory variable, with which the other series were

Table 3 Regression analysis of the calculated reaction energies

	Reaction series	<i>R</i>		<i>s</i>		F-test	α
		<i>meta</i>	<i>meta + para</i>	<i>meta</i>	<i>meta + para</i>		
<i>Set A (ionizations)</i>							
1	ArCH ₂ COOH	0.987	0.990	2.40	2.26	0.79	—
2	ArCH ₂ OH	0.9936	0.9923	2.20	2.43	1.42	—
3	ArC≡NH ⁺	0.975	0.966	4.57	6.65	3.11	0.10
4	ArCO ₂ H ₂ ⁺	0.971	0.965	5.36	7.36	2.67	0.10
5	ArNO ₂ H ⁺	0.971	0.941	5.74	10.96	5.94	0.01
6	ArC≡CH	0.9996	0.983	0.59	4.73	122.5	<< 0.005
7	ArOH	0.996	0.980	2.42	6.52	12.8	<< 0.005
8	Altogether			3.77	6.47	4.67	< 0.005
<i>Set B (uncharged molecules)</i>							
9	ArCO ₂ H	0.982	0.940	0.63	1.73	13.2	<< 0.005
10	ArC≡N	0.9912	0.982	0.83	1.30	3.74	0.05
11	ArNO ₂	0.9900	0.977	0.99	1.76	5.07	0.025
12	ArC≡CH	0.9928	0.910	0.27	0.96	23.8	<< 0.005
13	ArNH ₂	0.935	0.804	0.48	3.78	116.1	<< 0.005
14	ArN(CH ₃) ₂	0.976	0.865	0.59	3.79	77.0	<< 0.005
15	ArSH	0.991	0.097	0.31	1.93	73.0	<< 0.005
16	Altogether			0.63	2.42	26.7	<< 0.005

successively correlated. The results of regression are given in Table 3. In the set **A**, the *para*-substituents made the fit worse (very significantly) in most cases but not in all. Therefore, it would be possible to select a series of reactions without conjugation in which all substituents would be acceptable but the range of validity would be too restricted. Note that in all but one case the fit with the *para*-substituents would be still classified^{4a} as 'good fit', or still better.

In the set **B**, the differences between *meta*- and *para*-substituents are more marked. The good fit for *meta*-derivatives is remarkable taking into account that the range of validity has been exceeded. Note that in this set we are dealing with finer effects: compared to set **A** the substituent effects are smaller by an order of magnitude.

Principal component analysis (PCA) was carried out with non-standardized data (in the dimension kJ mol⁻¹) as in the previous work,⁷ otherwise the series with small values would be badly correlated and would influence the overall results too much. Three different arrangements were tested: (a) *meta*- and *para*-derivatives separately, (b) *meta*- and *para*-matrices merged vertically, each substituent taken twice as 3-X and 4-X as in the standard Hammett conception (denoted *meta/para* in Table 4), (c) *meta*- and *para*-matrices merged horizontally, each substituent taken only once, each reaction divided into *meta* and *para* as in the DSP treatment^{8b} (in Table 4 denoted *meta|para*). In addition to common PCA, we carried out all calculations also with the mathematical variant called the method of conjugated deviations.³⁵ Its merit is that it automatically eliminates the little significant series. In our case the results were practically not different and are not separately reported.

Results of PCA are given in Table 4. All components given are statistically significant but the ultimate ones were not interpreted and probably have no physical meaning. The results are in accord with those obtained from the experimental gas-phase acidities.⁷ The fit for *meta*-derivatives is much better: with one component it is so good as for *para*-deriva-

tives with two components (lines 1 and 2, 5 and 6). The original arrangement of Hammett (lines 3 and 7) is superior to that of DSP (lines 4 and 8) when the same number of two components is used. The best arrangement, however, is separate treatment of *meta*- and *para*-derivatives at different levels of complexity: one component is sufficient for *meta*-derivatives and two components are needed for *para*-. In the set **B** we got smaller residual deviations than in set **A** (because the energy values are smaller) but the percents of explained variability are similar.

Characterization of substituents

PCA gives the values of components, t_1, t_2, \dots, t_n , that can be used to characterize the substituents, reveal their mutual similarity, separation into classes,^{36,37} etc. Most important results were obtained with set **A** for the matrix *meta/para* corresponding to the standard Hammett equation (Table 4, line 3). The first component t_1 is closely correlated to the standard values^{8d} $\sigma_{m,p}$ as anticipated ($R = 0.983$). The second

Table 4 Principal component analysis of the calculated reaction energies^a

Source matrix	Number of components		
	1	2	3
<i>Set A (ionizations)</i>			
1 <i>meta</i>	3.10 [98.3]	1.78 [99.6]	
2 <i>para</i>	6.85 [95.9]	2.41 [99.7]	1.30 [99.95]
3 <i>meta/para</i>	5.72 [95.9]	2.66 [99.3]	1.50 [99.8]
4 <i>meta para</i>	5.94 [95.5]	3.71 [98.5]	1.77 [99.7]
<i>Set B (uncharged molecules)</i>			
5 <i>meta</i>	0.41 [99.2]	0.26 [99.7]	
6 <i>para</i>	1.95 [94.0]	0.94 [99.1]	0.46 [99.9]
7 <i>meta/para</i>	2.09 [89.3]	0.72 [99.0]	0.46 [99.7]
8 <i>meta para</i>	1.84 [91.5]	0.96 [98.1]	0.61 [99.4]

^a Residual standard deviation in kJ mol⁻¹, in square brackets percent of explained variability.

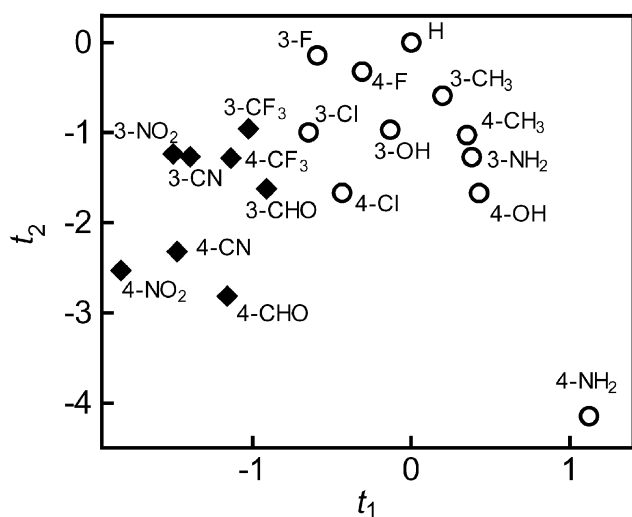


Fig. 4 Plot of the two first components of PCA treatment of ionization reactions (line 3 of Table 4): ○ donor substituents and hydrogen, ◆ acceptor substituents.

component t_2 expresses the enhanced resonance and has some resemblance to the constants σ_R^+ or σ_R^- . This is seen from Fig. 4. The components were plotted against each other, so that similar substituents are situated in the proximity. Separation of acceptors and donors is evident but clustering is not so strong as found earlier³⁶ with mixed experimental data also including many spectroscopic results. Our results agree merely with Wepster's description.³⁷ There is a fundamental difference between acceptors, more similar to each other, and donors with more graduating properties. We can confirm this and add that the substituent 4-NH₂ is unique. In the set **B** (Table 4, line 7) the values of components are somewhat different (Fig. 5). The acceptors form now a more differentiated group and behaviour of 4-CHO is unique. All *meta*-substituents are situated along a straight line; hence for them one component would be sufficient. In summary, clustering of substituents is in no case so strong that it could depreciate the standard statistical treatment. More consequential could be clustering of loadings of individual reactions. In the set **A**, three reactions, acidities of ArCOOH, ArCH₂COOH and ArCH₂OH, are too similar to each other, hence less efficient in the statistical procedures. In the set **B** the situation is better.

We searched still for possible clustering of substituents using the approach of the cluster analysis (CA) and obtained concordant results with the methods of furthest neighbour, average linkage and MacNaughton-Smith,³⁸ while the method nearest neighbour³⁸ was of little efficiency due to the chaining effect.³⁸ These results also essentially agreed with the plots of components like Fig. 4 and 5. For instance in the set **A**, CA gave five clusters: (H, 3-CH₃, 3-NH₂, 4-CH₃, 4-OH), (3-OH, 4-F, 3-F, 3-Cl, 4-Cl), (3-CF₃, 4-CF₃, 3-CHO, 4-CHO), (3-CN, 3-NO₂, 4-CN, 4-NO₂) and (4-NH₂). The pertinent tree diagram is reproduced as Fig. S1 (ESI†). In Fig. 4 these sets can be seen but they do not always form evident clusters but merely they appear as successions.

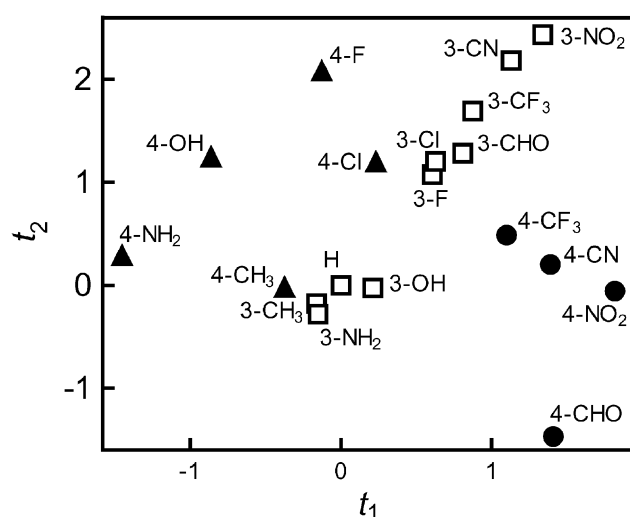


Fig. 5 Plot of the two first components of PCA treatment of reactions of uncharged groups (line 7 of Table 4): □ *meta*-substituents and hydrogen, ▲ *para* donor substituents, ● *para* acceptor substituents.

In our opinion, the sets of substituents commonly used are convenient for a statistical treatment since they cover satisfactorily the combinations of possible properties (essentially due to inductive and resonance effects). Our set was completely satisfactory in Fig. 5; in Fig. 4 perhaps an additional donor substituent would be helpful.

Conclusions

In several examples, quantum chemical calculations provided values of energies almost equally dependable as the values from experiments but broader and more complete. Their dependability was confirmed when part of the data was directly checked by comparison with experiments; this was done in all examples treated in this work.^{17,23–26} In our opinion, there is no doubt that quantum chemical calculations have contributed significantly to the theory of substituent effects and to structure–reactivity relationships.

In the case of the Hammett equation, our calculations gave results concordant with the previous treatment of experimental gas-phase data⁷ and with many observations made with the individual conjugated substituents in solution. However, the quantum chemical calculations gave more systematic and more convincing results, some of them are hardly obtainable from experiments. Most striking is the finding that the Hammett equation is valid even for reactions of uncharged particles and to reactions proceeding directly on the benzene ring, but only for *meta*-derivatives. There is merely a formal problem, which conclusion one wants to draw. We suggest that the range of validity of the Hammett equation should be restricted strictly to *meta*-derivatives, in the theoretical studies and in the textbooks. On the other hand, we know that the original, broader range of validity will be still used when interpreting scarce experimental data but care must be taken to eliminate all cases of direct conjugation.

Acknowledgements

This work was supported by the projects Z4 055 0506 of the Academy of Sciences (to O. E.) and MSM0021627501 of the Ministry of Education, Youth and Sports of the Czech Republic (to P. P.).

References

- 1 L. P. Hammett, *Physical Organic Chemistry*, McGraw-Hill, New York, 2nd edn, 1970, ch. 11.9.
- 2 (a) O. Exner, *Correlation Analysis of Chemical Data*, Plenum Press, New York, 1988, ch. 2; (b) O. Exner, in *Advances in Linear Free Energy Relationships*, ed. N. Chapman and J. Shorter, Plenum Press, London, 1972, pp. 1–69.
- 3 For instance: (a) E. C. Lee, B. H. Hong, J. Y. Lee, J. C. Kim, D. Kim, Y. Kim, P. Tarakeshwar and K. S. Kim, *J. Am. Chem. Soc.*, 2005, **127**, 4530–4537; (b) N. H. Nguyen, J. W. Apriletti, J. D. Baxter and T. S. Scanlan, *J. Am. Chem. Soc.*, 2005, **127**, 4599–4608.
- 4 (a) H. H. Jaffé, *Chem. Rev.*, 1953, **53**, 191–261; (b) H. van Bekkum, P. E. Verkade and B. M. Wepster, *Recl. Trav. Chim. Pays-Bas*, 1959, **78**, 815–850; (c) Y. Okamoto and H. C. Brown, *J. Org. Chem.*, 1957, **22**, 485–494; (d) M. Sjöström and S. Wold, *Chem. Scr.*, 1976, **9**, 200–210.
- 5 (a) C. Hansch, A. Leo and R. W. Taft, *Chem. Rev.*, 1991, **91**, 165–195; (b) O. Exner, in *Correlation Analysis in Chemistry: Recent Advances*, ed. N. B. Chapman and J. Shorter, Plenum Press, New York, 1978, pp. 439–540.
- 6 (a) H. H. Jaffé, *J. Am. Chem. Soc.*, 1959, **81**, 3020–3021; (b) J. Hine, *J. Am. Chem. Soc.*, 1959, **81**, 1126–1129.
- 7 M. Ludwig, S. Wold and O. Exner, *Acta Chem., Scand.*, 1992, **46**, 549–554.
- 8 (a) Y. Yukawa, Y. Tsuno and M. Sawada, *Bull. Chem. Soc. Jpn.*, 1966, **39**, 2274–2286; (b) S. Ehrenson, R. T. C. Brownlee and R. W. Taft, *Prog. Phys. Org. Chem.*, 1973, **10**, 1–80; (c) M. Charton, *Prog. Phys. Org. Chem.*, 1981, **13**, 119–251; (d) M. Charton, *Prog. Phys. Org. Chem.*, 1987, **18**, 287–315; (e) G. Weeks and V. Horák, *J. Org. Chem.*, 1980, **45**, 2068–2071; (f) G. H. E. Nieuwdorp, C. L. de Ligny and H. C. van Houwelingen, *J. Chem. Soc., Perkin Trans. 2*, 1979, 537–544.
- 9 J. Shorter, *Correlation Analysis of Organic Reactivity with Particular Reference to Multiple Regression*, Research Studies Press, John Wiley & Sons, Chichester, 1982.
- 10 (a) O. Pytela, *Collect. Czech. Chem. Commun.*, 1995, **60**, 1502–1528; (b) O. Pytela, *Collect. Czech. Chem. Commun.*, 1996, **61**, 704–712.
- 11 J. C. R. Reis, M. A. P. Segurado and J. D. G. de Oliveira, *J. Phys. Org. Chem.*, 1995, **8**, 671–688.
- 12 (a) O. Exner, *Prog. Phys. Org. Chem.*, 1990, **18**, 129–161; (b) K. Palát Jr, S. Böhm, G. Braunerová, K. Waisser and O. Exner, *New J. Chem.*, 2002, **26**, 861–866.
- 13 K.-C. Zhang, Y. Fu, B.-L. Lin, L. Liu, Y.-H. Cheng and Q.-X. Guo, *J. Mol. Struct. (THEOCHEM)*, 2003, **625**, 199–205.
- 14 (a) J. Shorter, *Pure Appl. Chem.*, 1994, **66**, 2451–2468; (b) J. Shorter, *Pure Appl. Chem.*, 1997, **69**, 2497–2510.
- 15 (a) T. B. McMahon and P. Kebarle, *J. Am. Chem. Soc.*, 1977, **99**, 2222–2230; (b) R. W. Taft and R. D. Topsom, *Prog. Phys. Org. Chem.*, 1987, **16**, 1–83.
- 16 (a) S. Böhm and J. Kuthan, *Int. J. Quantum Chem.*, 1984, **26**, 21–33; (b) O. Exner, D. Ondřejíková and J. Leška, *Org. React. (Tartu)*, 1979, **16**, 285–295; (c) T. Sotomatsu, Y. Murata and T. Fujita, *J. Comput. Chem.*, 1989, **10**, 94–98; (d) L. Liu, Y. Fu, R. Liu, R.-Q. Li and Q.-X. Guo, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 652–657.
- 17 O. Exner and S. Böhm, *J. Org. Chem.*, 2002, **67**, 6320–6327.
- 18 K. B. Wiberg, *J. Org. Chem.*, 2002, **67**, 4787–4794.
- 19 (a) A. Pross, L. Radom and R. W. Taft, *J. Org. Chem.*, 1980, **45**, 818–826; (b) O. Exner, *Org. React. (Tartu)*, 1995, **29**, 1–6.
- 20 P. George, M. Trachtman, C. W. Bock and A. M. Brett, *J. Chem. Soc., Perkin Trans. 2*, 1976, 1222–1227.
- 21 O. Exner and S. Böhm, *Collect. Czech. Chem. Commun.*, 2001, **66**, 1623–1637.
- 22 Y. Fu, L. Liu, R.-Q. Li, R. Liu and Q.-X. Guo, *J. Am. Chem. Soc.*, 2004, **126**, 814–822.
- 23 K. B. Wiberg, *J. Org. Chem.*, 2003, **68**, 875–882.
- 24 S. Böhm and O. Exner, *New J. Chem.*, 2005, **29**, 336–342.
- 25 O. Exner and S. Böhm, *Phys. Chem. Chem. Phys.*, 2004, **6**, 3864–3871.
- 26 O. Exner and S. Böhm, *Org. Biomol. Chem.*, 2005, **3**, 1838–1843.
- 27 K. B. Wiberg, *Collect. Czech. Chem. Commun.*, 2004, **69**, 2183–2192.
- 28 (a) A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100; (b) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789; (c) B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett.*, 1989, **157**, 200–206; (d) A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 29 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision B.03)*, Gaussian, Inc., Pittsburgh, PA, 2003.
- 30 J. Roithová and O. Exner, *J. Phys. Org. Chem.*, 2001, **14**, 752–758.
- 31 (a) K. B. Wiberg and K. E. Laidig, *J. Am. Chem. Soc.*, 1987, **109**, 5935–5943; (b) R. F. W. Bader, J. R. Cheeseman, K. E. Laidig, K. B. Wiberg and C. Breneman, *J. Am. Chem. Soc.*, 1990, **112**, 6530–6536.
- 32 S. Böhm and O. Exner, *Org. Biomol. Chem.*, 2003, **1**, 1176–1180.
- 33 S. Böhm and O. Exner, *J. Mol. Struct. (THEOCHEM)*, 2005, **722**, 125–131.
- 34 S. Böhm, J.-F. Gal, P.-C. Maria, J. Kulhánek and O. Exner, *Eur. J. Org. Chem.*, 2005, 2580–2588.
- 35 O. Pytela, *Collect. Czech. Chem. Commun.*, 1990, **55**, 634–643, and papers cited therein.
- 36 (a) D. Johnels, U. Edlund, H. Grahn, S. Hellberg, M. Sjöström, S. Wold, S. Clementi and W. J. Dunn, *J. Chem. Soc., Perkin Trans. 2*, 1983, 863–871; (b) S. Alunni, S. Clementi, C. Ebert, P. Linda, G. Musumarra, U. Edlund, M. Sjöström and S. Wold, *Gazz. Chim. Ital.*, 1986, **116**, 679–682.
- 37 A. J. Hoefnagel, W. Oosterbeek and B. M. Wepster, *J. Org. Chem.*, 1984, **49**, 1993–1997.
- 38 P. MacNaughton-Smith, *Some Statistical and Other Numerical Techniques for Classifying Individuals*, Home Office Research Unit Report, London, 1965.