ARTICLE

OBC
www.rsc.org/obc

Quantum chemical topology (QCT) descriptors as substitutes for appropriate Hammett constants

P. J. Smith and P. L. A. Popelier*

School of Chemistry, Faraday Building, Sackville Site, University of Manchester, Manchester, M60 1QD, Great Britain. E-mail: pla@manchester.ac.uk; Fax: +44 (0)161 306 4559; Tel: +44 (0)161 306 4511

Received 18th May 2005, Accepted 26th July 2005
First published as an Advance Article on the web 17th August 2005

A technique called quantum topological molecular similarity (QTMS) was recently proposed [*J. Chem. Inf. Comput. Sci.*, 2001, **41**, 764] in order to construct a variety of medicinal, ecological and physical organic QSAR/QSPRs, based on modern *ab initio* wave functions of geometry optimised molecules, in combination with quantum chemical topology (QCT). The current abundance of computing power can be utilised to inject realistic descriptors into QSAR/QSPRs. In previous work [*J. Chem. Soc., Perkin Trans. 2*, 2002, 1231] it was proven that a set of Hammett constants (σ_p, σ_m, σ_I and σ_p^0) for a sizeable set of mono- and polysubstituted carboxylic acids can be replaced by QCT bond descriptors. Using QTMS and proper statistical validation we examined seven data sets in total. The first three sets (*para*-substituted phenols (σ^-), substituted toluenes (σ^+) and bromophenethylamines (σ^+)) corroborate that a wider class of Hammett constants can also be replaced by QCT descriptors. A fourth set (benzyl radicals) focuses on non-Hammett behaviour being superimposed on Hammett behaviour. QCT descriptors selectively correlate with Hammett behaviour. The QTMS analysis of the last three sets (toxicity of benzyl alcohols, chromatographic capacity factors of chalcones and herbicidal activity of 5-chloro-2,3-dicyanopyrazines) screens for false positives. This test is successfully passed in that QCT descriptors fail when lipophilicity/hydrophobicity is in charge. Hence, overall, the discriminatory capacity of QCT descriptors is established, in detecting Hammett behaviour and specifically replacing the Hammett constants by more modern and non-empirical descriptors.

Introduction

The Hammett substituent constants, which quantify the electron withdrawing or donating capabilities of a given substituent, have enjoyed enormous success as structural descriptors in quantitative structure–activity relationships (QSARs) and quantitative structure–property relationships (QSPRs).[1] The physical chemical database of Hansch and co-workers contains over 8900 QSARs, of which 8875 are based on the σ parameters.[2] It is apparent that electronic effects are of considerable importance in describing chemical reactions, and hence biological systems.

Some time ago[3] we used for the first time quantum chemical descriptors defined by quantum chemical topology (QCT)[4–7] to set up a QSAR to predict σ constants and, therefore, also the acidity of substituted benzoic acids. Other than being a generalisation of quantum mechanics to subspaces[8–10] QCT is widely appreciated as a theory that extracts chemical insight from modern wave functions.[11,12] The acronym QTMS (quantum topological molecular similarity) covered the initial development and intention of QCT to provide an economic alternative to Carbo-like indices.[13] It was proven that superposition of (complete molecular) electron densities is not necessary to set up a successful QSAR. Instead, as explained below, special points in 3D space sufficed.

Although QCT descriptors could have been fed into molecular similarity indices, they immediately appeared in a "supervised" regression context. In other words, models[14] were constructed using partial least squares (PLS) for a variety of medicinal[15–17] and ecological[18–21] QSARs. The estimation of the pK_a of carboxylic acids, anilines and phenols[22] and prediction of σ_p, σ_m, σ_I and σ_p^0 parameters of mono-[3] and polysubstituted benzoic acids, phenylacetic acids and bicyclo carboxylic acids[19] feature as examples of physical organic properties. The QTMS approach inspired work in other groups (*e.g.*ref. 23–24). A strong and important feature of QTMS is that it is able to localise a part in the molecule where the chemical change associated with the observed activity actually happens. For example, the O–H bond is highlighted as the active center in the deprotonation of carboxylic acids if their acidity is studied.

QTMS typically uses so-called bond critical point (BCP) properties, as in this study, but (integrated) atomic properties can also feature. BCPs are 3D saddle points in the electron density, located by computation[25] and appearing at the boundary between two QCT atoms. Certain functions, such as the electron density, are evaluated at each BCP, thereby characterising the bond that the BCP represents. A BCP can be represented by an arbitrarily high number of properties, which serve as vector components locating the BCP as a "quantum chemical fingerprint" in BCP-space.[3]

In view of the ubiquity of the Hammett parameters in formulating QSARs, coupled with the promising results obtained so far in capturing substituent effects, we aimed at extending the range of σ constants investigated by QTMS. In addition, the range of systems is hereby also increased, beyond the domain of carboxylic acid datasets. We examined four data sets: *para*-substituted phenols (σ^-), bromophenethylamines (σ^+), substituted toluenes (σ^+), and benzyl radicals in order to investigate non-Hammett behaviour. We will show that excellent statistics are obtained for the first three sets, and that the active centers highlighted make sense. The last set is a "false positive" test, adopted to see whether QTMS still performs well when conventional "polar effect" Hammett constants fail. We will show that QTMS passes this test and hence operates successfully for the right reasons.

Methods

The full details of how QTMS operates have been published before[13] and repeated in the applications mentioned in the Introduction. In the interest of making this paper self-contained it is useful to summarise the procedure here.

Geometry-optimised wave functions were generated at HF/6-31G*//HF/6-31G* level using the program GAUSSIAN98.[26] In previous QTMS work we varied the level of theory between

DOI: 10.1039/b507024d

AM1 and B3LYP/6-311 + G(2d,p)//B3LYP/6-311 + G(2d,p), which obviously lead to substantial differences in consumed CPU time. In our experience, HF/6–31G*//HF/6–31G* is a good compromise in terms of speed and reliability. As shown in the Results section, excellent statistical correlations are obtained at this level, directly corroborating this choice. This success suggests that the approximations this level introduces generate systematic errors, which are largely absorbed by the differential nature of a QSAR. Ultimately, we are only interested in the relative differences that substituents bring in.

The program MORPHY98[27] computes the following four BCP properties for each BCP in each molecule of the congeneric set of compounds: the electron density, the Laplacian of the electron density, the ellipticity and the kinetic energy density $K(\mathbf{r})$. The definition and meaning of these quantities has been given elsewhere.[7,16]

Subsequently the program SIMCA-P[28] performs a PLS[29] analysis directly on the BCP descriptors. Each latent variable (LV) is created from a linear combination of the original independent variables and in this respect LVs are similar to principal components[30] (but now in a supervised context). This analysis yields four statistics, which are used to judge the quality and validity of the model. The first two statistics provide a measure of the quality of the model and are the correlation coefficient, r^2, and the cross-validated correlation coefficient, q^2, which we will explicitly quote for each QSAR set. The final two statistics obtained provide a safeguard against the possibility that the model may have been obtained by chance: the response data is repeatedly randomised and a re-run (at least 10 times) of the PLS analysis carried out. We do not quote these statistics but confirm that according to the prescription of Wold and co-workers[29] all QSAR sets studied here pass this test.[31]

In order to see if the reaction center can be located, the program SPSS[32] extracted PCs from the BCP properties associated with each bond and a second PLS regression was performed, now operating on the PCs. This time we examine the variables important to the projection (VIPs), which are also generated by SIMCA-P. The VIPs give the relative importance of each PC contributing to the model, and PCs with higher VIP scores are more relevant in explaining the activity. Since each PC is associated with a particular bond this allows one to rank the importance of each bond contributing to the model. It is assumed that the "active center" of the molecules consists of the bonds associated with the PCs having the highest VIP values. To summarise, the first PLS analysis yields the regression statistics, which assess the quality and validity of the model, whereas the second PLS analysis focuses upon the VIP values of the PCs in order to assist interpretation of the mode of action of the QSAR.

In addition, a genetic algorithm (GA) selected individual BCP properties to be included as descriptor variables in a third PLS analysis. The purpose of this was twofold: to optimise the regression statistics, and to see if the GA can be used to locate the active center. Variable selection precedes the PLS fitting but occurs immediately after the BCP properties have been calculated. The MATLAB *genalg.m* routine from the PLS_toolbox[33] selected the variables, using a GA population size of 256, a mutation rate of 0.003, and a maximum number of generations of 200. The fitness function was defined to be the cross-validation coefficient, q^2, from a PLS analysis performed on the dataset using variables selected by the GA.

Results and discussion

para-Substituted phenols and σ^-

The first set studied demonstrates how QTMS can be extended to incorporate through-resonance effects. Despite the enormous success of the original Hammett equation, deviations were observed for systems where the reaction center is in direct conjugation with substituents capable of accepting or donating electrons.[34] This behaviour is exemplified in the ionisation of phenols that are bonded to substituents such as NO_2, as shown in Fig. 1. This system is inadequately modelled by σ, because the original constants were obtained from benzoic acids, which do not exhibit through-resonance, and hence the two systems are not comparable. In order to overcome these failings, a new set of parameters, known as σ^-, were introduced. The values of these constants are derived from anilines and phenols, and differ from the conventional Hammett constants for electron accepting substituents (for example NO_2). For other substituents that cannot accept electrons, such as halogens and alkyl groups for example, the differences are less pronounced. It should also be noted that these constants only apply to substituents in the *para* position. Strictly speaking, the σ^- scale was not introduced nor defined for *meta* substituents and hence both sets of values, σ^- and σ, are exactly the same.

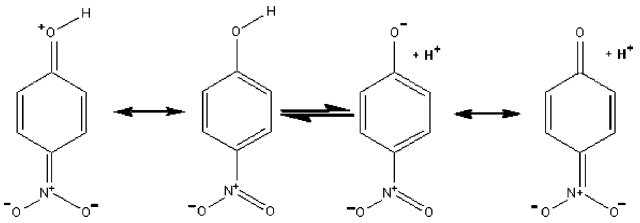

Fig. 1 Resonance structures for *p*-nitrophenol and the *p*-nitrophenol anion.

We have performed a QTMS analysis on a set of 17 *para*-substituted phenol molecules in order to gauge the ability of the method to reproduce σ^- values. The structure and substituent constants for the compounds are given in Table 1. The common molecular skeleton is shown in Fig. 2. Sigma values were obtained from ref. 34 and have been chosen to reflect a range of electron-accepting/donating abilities. The PLS analysis returns $r^2 = 0.983$ and $q^2 = 0.938$ with 3 LVs. An excellent model is obtained, confirming the ability of our method to reproduce Hammett constants for sets of compounds where there is direct conjugation between the substituent and reaction center. The model validates according to the randomisation test. An observed *versus* predicted plot for the σ^- values is shown in Fig. 3, illustrating the range and spread of the data.

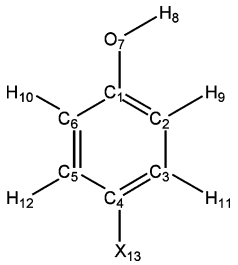

Fig. 2 Numbered common skeleton of the *p*-phenol molecules.

In conjunction with the GA, the PLS analysis returns improved correlation statistics: $r^2 = 0.993$ and $q^2 = 0.969$, now with 4 LVs. The corresponding VIP plots, with and without the

Table 1 Substituents and σ^- values for the set of substituted phenols

Molecule	Substituent	σ^-	Molecule	Substituent	σ^-
1	H	0	10	CN	1
2	OH	−0.37	11	NO_2	1.27
3	OCH_3	−0.26	12	NH_2	−0.15
4	$C(CH_3)_3$	−0.13	13	CF_3	0.65
5	CH_3	−0.17	14	$CONH_2$	0.61
6	C_6H_5	0.02	15	CCH	0.53
7	F	−0.03	16	$CH(CH_3)_2$	−0.16
8	Cl	0.19	17	$COOCH_3$	0.64
9	Br	0.25			

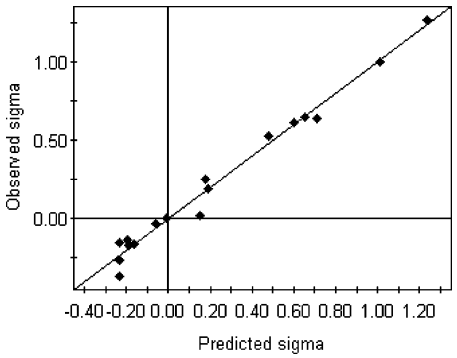

Fig. 3 Observed *versus* predicted σ^- values for *p*-substituted phenols.

GA, are given in Fig. 4a and b, respectively. It can be seen that the GA greatly reduces the number of independent variables, from 52 (= 4 × 13) to 11 (Fig. 4a). In terms of the active center, both procedures (the GA selecting BCP properties and the PCs without GA involvement) highlight the OH bond (bond 7–8). This is the bond one would expect to be the most important in explaining the activity. Unfortunately, however, both methods suffer from "contaminations" since variables with high VIP scores cannot be directly related to the activity. Two variables associated with bond 3–11 (a CH bond) are selected by the GA and have a high VIP score; this bond is also given high priority in the VIP plot for the PCs. It is difficult to reconcile these bonds with the mechanism for phenol dissociation, although it is pleasing to note that variables associated with the correct bond do come top in both VIP plots. We note that in their study on the identification of active molecular sites using quantum-self-similarity measures Amat *et al.*[35] also detected unexpected fragments correlating strongly with σ constants. These authors suggested a possible explanation for this phenomenon based on Mezey's "holographic electron density theorem".[36]

Substituted toluenes and σ^+

In addition to the σ^- constants, there also exist analogous σ^+ constants, which account for substituents with the ability of delocalising a positive charge. An example of where the σ^+ constants have found considerable success is in describing electrophilic aromatic substitution and electrophilic side-chain reactions. In these reactions the stability of the transition state, and hence reaction rate, is greatly influenced by through-resonance and therefore unsatisfactorily correlated by the Hammett equation. This problem was addressed by Brown and Okamoto,[37] who successfully correlated rate data for electrophilic aromatic substitution with σ^+ constants derived from the solvolysis of cumyl chlorides in 90% aqueous acetone. This has proven to be an ideal model for systems capable of delocalising a positive charge.

Table 2 Activity and σ^+ values for the set of substituted toluenes

Molecule	X[a]	Y[a]	$\sigma^{+b,c}$	Activity[d]	log (activity)
1	H	OCH_3	−0.778	11.7	1.068
2	H	CH_3	−0.311	2.56	0.408
3	H	t-C_4H_9	−0.256	2	0.301
4	CH_3	H	−0.066	1.52	0.182
5	H	H	0	1	0.000
6	OCH_3	H	0.047	0.75	−0.125
7	H	Br	0.15	0.94	−0.027
8	H	Cl	0.114	0.8	−0.097
9	H	F	−0.073	0.58	−0.237
10	COOH	H	0.322	0.3	−0.523
11	H	COOH	0.421	0.22	−0.658
12	Br	H	0.405	0.24	−0.620
13	H	CN	0.659	0.14	−0.854
14	CN	H	0.562	0.11	−0.959
15	NO_2	H	0.674	0.08	−1.097
16	H	NO_2	0.79	0.05	−1.301

[a] Fig. 8: X = *meta*,Y = *para*. [b] This sigma value is not a sum, but simply refers to either the *meta* or *para* substituent. [c] Note that the value for the *para*-chloro substituted molecule is slightly different to the corresponding entry in Table 3. This is due to slight discrepancies in the experimental papers from which the current values were drawn. Although the σ^+ values should be the same in both tables the small differences do not alter our findings. [d] Relative reactivities of substituted toluenes toward *N*-bromosuccinimide (per hydrogen) at 80 °C in CCl_4.

We have chosen to analyse a set of 16 *meta* and *para*-substituted toluenes and their relative reactivities towards *N*-bromosuccinimide, a bromination reaction that is well described using σ^+. The reaction rate data (relative to toluene itself) were taken from Walling *et al.*[38] The substituent constant values were obtained from Brown and Okamoto[37] and are presented in Table 2 (note that one of the substituents is always a hydrogen). The common skeleton of the toluenes is shown in Fig. 5.

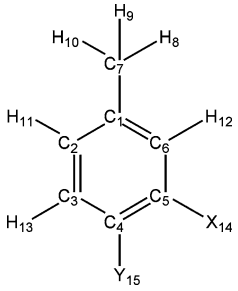

Fig. 5 Numbered common skeleton for the substituted toluenes.

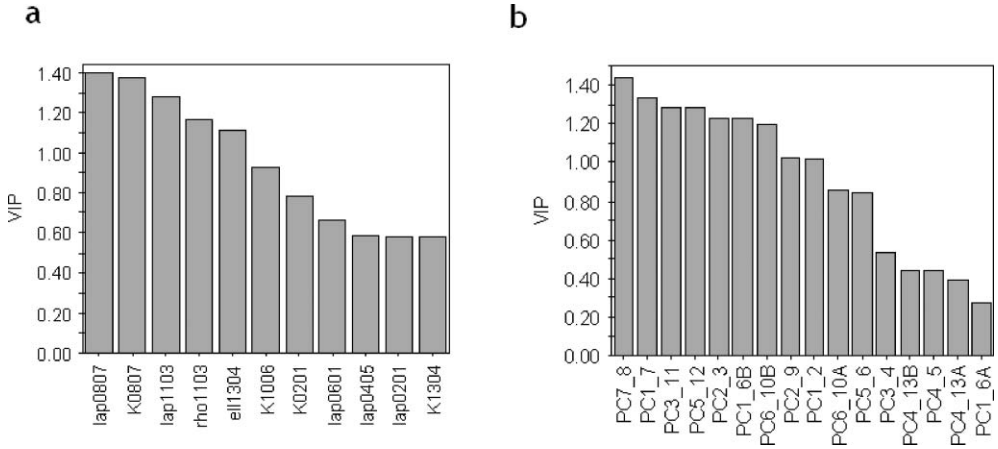

Fig. 4 PLS analysis on *para*-substituted phenols. VIP values of descriptors: (a) BCP properties selected by the GA and (b) principal components (without GA selection).

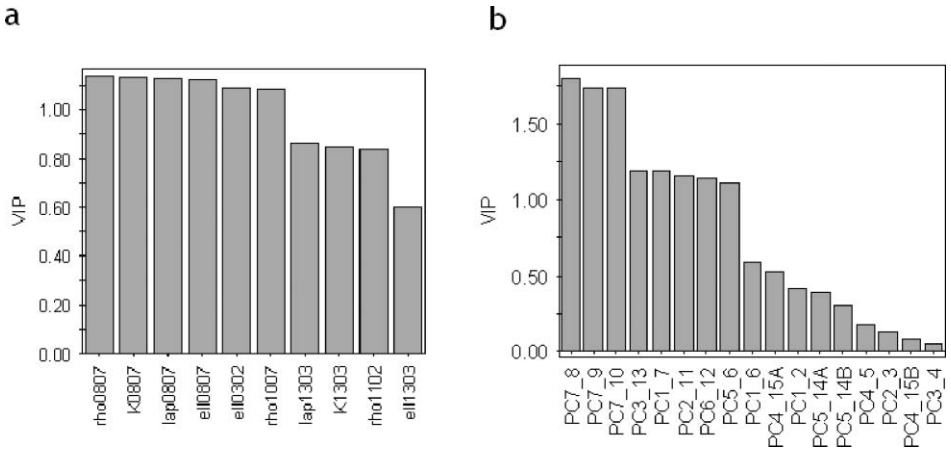

Fig. 6 PLS analysis on substituted toluenes. VIP values of descriptors: (a) BCP properties selected by the GA and (b) principal components (without GA selection).

The results of the PLS regression (randomization validated) using all the BCP properties as descriptors, and the logarithm of the relative rate of reactivity towards *N*-bromosuccinimide as the dependent variable, are $r^2 = 0.917$ and $q^2 = 0.817$ for only 1 LV. This is reasonable but not impressive considering that regressing the reaction rate with σ^+ alone produces a model with $r^2 = 0.955$. In order to optimise the PLS model, a GA was employed to select "successful" BCP properties as descriptors. PCs were also extracted from the BCP properties and used as descriptor variables in order to isolate the active center. The fitness function was taken to be the cross-validation error (q^2) from a PLS analysis performed on the dataset using variables selected by the GA. The regression statistics of the PLS model fitted using the GA-selected BCP properties improved considerably to $r^2 = 0.958$ and $q^2 = 0.891$ with 2 LVs. The corresponding VIP plots, with and without the GA, are given in Fig. 6a and b, respectively.

The model fitted using the GA selected BCP properties is superior to the one obtained using all the properties, with the predictivity in particular showing a marked improvement. The statistics obtained are similar to those quoted by Walling *et al.*[38] and confirm the ability of QTMS to successfully model a dataset whose activities exhibit good correlations with another kind of Hammett constant, σ^+. Indeed, fitting another PLS model using the same GA-selected BCP properties as descriptors, regressed against σ^+ itself, yields an excellent correlation with $r^2 = 0.973$ and $q^2 = 0.930$.

With regard to the active center in the molecules, both methods agree that bond C_7H_8 (in the methyl group) is the most important in producing good correlations. This makes perfect sense, since one of the hydrogens on the methyl group is abstracted in the formation of a $X–C_6–H_5CH_2^{\bullet}$ radical. Thus the C–H bond strength is crucial in determining the reaction rate, which is why QTMS highlights this bond. In the VIP plot for the variables selected by the GA (Fig. 6a) the top four variables are all associated with the same bond, namely C_7H_8. The VIP plot of the PCs (Fig. 6b) also places the C_7H_8 bond top, along with the two other CH bonds in the methyl group. This is to be expected since all three bonds are nearly identical in terms of BCP property values, and the trends exhibited by the properties. The construction of PCs guarantees a more viable picture of the active center than one based on a GA selected set of BCP properties. It is still reassuring to note that there is a large discrimination between the three bonds in the methyl group and the remainder of the bonds in the toluene molecules. The methyl group can be considered the true reaction center.

We just note here that an even sharper decline was found in a VIP plot[39] of BCP properties (without GA selection) in a set of 68 substituted carboxylic acids, prominently highlighting the O–H and C–O bonds in the COOH group ($r^2 = 0.926$ and $q^2 = 0.904$), as one would expect.

Bromophenethylamines and σ^+

To diversify the range of systems examined, we have looked at a set of 20 *meta*-, *para*-, and disubstituted *N*,*N*′-dimethyl-2-bromophenethylamines. The bromophenethylamines are believed to exhibit anti-adrenergic activity *via* the formation of benzylic cations, which the σ^+ constants are particularly well suited to describing. The Hammett σ^+ values were obtained from ref.[40] Fig. 7 shows the common skeleton for this second set of molecules and Table 3 contains the substituent data.

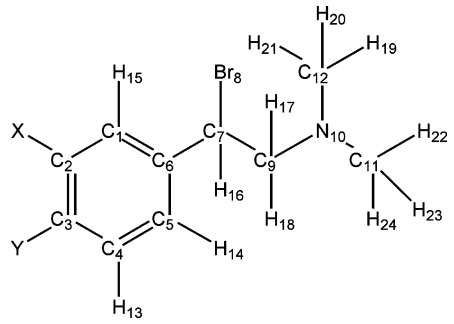

Fig. 7 Numbered common skeleton of the bromophenethylamines.

Again an excellent and valid model is produced, since $r^2 = 0.979$ and $q^2 = 0.951$ for 3 LVs. All the molecules are satisfactorily modelled, with no noticeable outliers, as can be seen from the observed *versus* predicted plot in Fig. 8.

Table 3 Substituents[a] (X and Y) and σ^{+b} values for the bromophenethylamines

Molecule	X	Y	σ^{+b}	Molecule	X	Y	σ^{+}
1	H	H	0	11	Br	F	0.34
2	H	F	−0.07	12	Me	F	−0.14
3	H	Cl	0.11	13	Cl	Cl	0.51
4	H	Br	0.15	14	Br	Cl	0.52
5	H	Me	−0.31	15	Me	Cl	0.04
6	F	H	0.35	16	Cl	Br	0.55
7	Cl	H	0.4	17	Br	Br	0.56
8	Br	H	0.41	18	Me	Br	0.08
9	Me	H	−0.07	19	Me	Me	−0.38
10	Cl	F	0.33	20	Br	Me	0.1

[a] Fig. 5: X = *meta*, Y = *para*. [b] The σ^+ is actually a sum of the σ_m^+ and σ_p^+ constants, as originally reported in ref. 41 However, it is more rigorous (and common in the LFER literature) to use the symbol $\Sigma\sigma$, since the σ^+ scale has actually not been introduced for *meta* substituents.

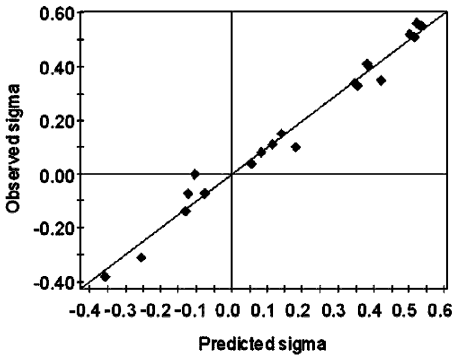

Fig. 8 Observed *versus* predicted σ^+ values for the bromophenethylamines.

Due to the impressive statistics obtained using all the BCP properties, it was deemed unnecessary to employ variable selection for this dataset. However, a second PLS analysis was performed on the molecules using PCs extracted from the BCP properties in an effort to locate the active center. The VIP plot for the PCs is shown in Fig. 9. It can be seen from the VIP plot that the highlighted region in the molecules, as defined by the VIP scores, is quite diffuse. Several bonds acquire similar scores, and there is no clear drop-off in the values for the first few descriptor variables. On the plus side, all the PCs with high scores are associated with bonds on the side-chain attached to the phenyl ring, which can be considered the reaction center for this set of molecules as it is this group that bears the positive charge upon formation of the benzylic cation.[41] This positive charge can be located on either the nitrogen attached to the two methyl groups or the carbon attached to the bromine atom.

Benzyl radicals and non-Hammett behaviour

One of the most fundamental concepts behind the Hammett equation is that of the electronic demand, which classifies each substituent either as an electron donor or as an electron acceptor. In other words, the substituent either supplies or removes electron density from some reaction center, thereby shifting the properties of a system in opposite directions. There are occasions, however, when both types of substituent shift the property of a system *in the same direction.*[42,43] This anomaly is known as non-Hammett behaviour. An example of this phenomenon was already noticed more than 50 years ago[44] in the context of the substituent effect on radical stability, although those authors did not explicitly use the term "non-Hammett behaviour". This term appeared a decade later, for example in Walter's work[45] on substituents affecting the properties of stable aromatic free radicals.

Given the success of QTMS in dealing with systems displaying ordinary Hammett behaviour, the next logical step is to see how the method performs with systems showing *non*-Hammett behaviour.

Liu *et al.*[43] have investigated the stabilities of 56 benzyl radicals exhibiting various degrees of non-Hammett behaviour. In this case the dependent variable is the reaction energy of the isodesmic reaction shown in eqn 1, where all 14 X substituents are always in the *para* position and Y = H, F, Cl or Li.

$$X\text{–}C_6H_4\text{–}CHY^{\bullet} + C_6H_6 \rightarrow C_6H_5\text{–}CHY^{\bullet} + X\text{–}C_6H_5 \qquad (1)$$

Liu *et al.* computed the reaction energies at the B3LYP/6-311++G(2d,2p)//B3LYP/6-31G(d) level with the B3LYP/6-31G(d) zero point energies scaled by 0.9806. For our purpose these reaction energies are adopted as independent observations (*i.e.* Y variables), just as measured reactivities featured in the set of bromophenethylamines. The authors correlated reaction energies with both conventional Hammett parameters and a variety of special scales of substituent constants used in radical chemistry ($\sigma^{\bullet}$). Here we focus on Jackson's σ_J constant,[46] which reflects carbon radical stability and which is suited to describing non-Hammett behaviour.

We have examined the same set of 56(= 4 × 14) benzyl radicals using QTMS in order to assess the ability of the method to deal with non-Hammett datasets. Table 4 contains the substituent data and reaction energies for the reaction shown in eqn 1. The common skeleton of the molecules is shown in Fig. 10. Table 5 shows the results obtained by Liu *et al.* correlating the

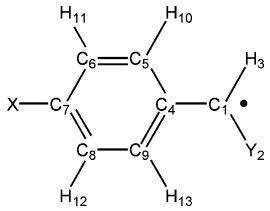

Fig. 10 Common skeleton and numbering scheme for the benzyl radicals.

Table 4 Substituent and activity data for the benzyl radicals

Molecule	Substituent	Energy change/kJ mol^{-1}			
	X	Y = H	Y = F	Y = Cl	Y = Li
1	CH_3	1.1	1.1	1.6	−1.2
2	Cl	1.2	0.2	0.3	5.4
3	CN	5	4.1	2.8	18.8
4	$COCH_3$	6.1	6	4.8	16.5
5	COOH	5.2	5.5	4.1	15.3
6	F	−0.9	−2.1	−1.7	1.2
7	H	0	0	0	0
8	NO_2	6.1	5.6	3.7	25.2
9	$CONH_2$	3.7	3.4	2.5	12.4
10	SCH_3	4.2	3.5	4.4	2.5
11	CF_3	1.1	0.5	−0.3	11.4
12	$N(CH_3)_2$	5.6	5	7.4	−3.7
13	COOMe	4.3	4.7	3.6	12.3
14	$SiMe_3$	2	2.1	2.2	2.5

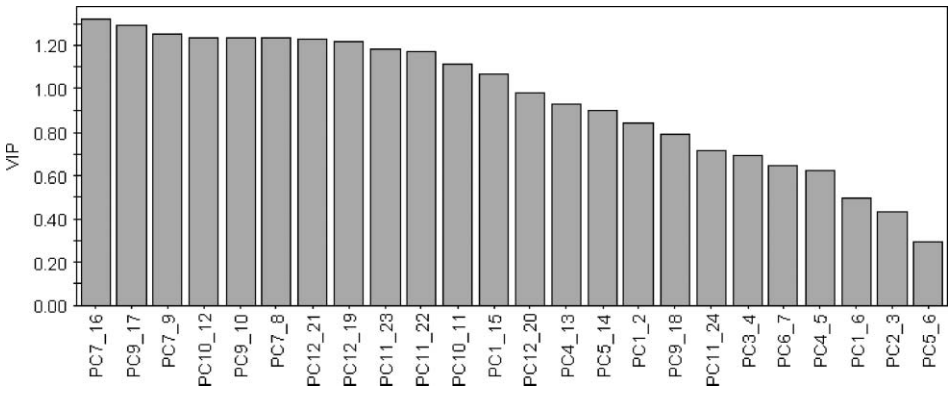

Fig. 9 PLS analysis on bromophenethylamines. VIP values of principal component descriptors (without GA selection).

Table 5 Regression correlation coefficients (r^2) of the energy changes of the isodesmic reaction (eqn 1) against different electronic descriptors

Y	σ_p^+ model	σ_J model	BCP model
H	0.0004[a]	0.9604	0.505
F	0.0001	0.9025	0.567
Cl	0.1936	0.8464	0.446
Li	0.6400[b]	0.5776	0.973

[a] The twelve entries in this table must not be confused with the sigma values themselves. [b] The r^2 value for the σ_p^- model is 0.9801.

reaction energies with the conventional Hammett σ_p^+ constant and with σ_J. Note that the Hammett constants always refer to the X substituents, not the Y substituents. Table 5 also lists the results from the PLS regressions using BCP properties as descriptors. Only r^2 values are included, since these are the only statistics included in the literature (*i.e.* no q^2 values are quoted). Again, all the models passed the validation test.

It can be seen that for all of the datasets, with the exception of the one containing lithium (Y = Li), the substituent effect on the reaction energy is poorly correlated by the conventional Hammett constant σ_p^+. However, the σ_J constants show the opposite trend, producing good fits for three of the sets (Y= H, F, Cl) but a poor correlation for the one containing lithium. It is evident that σ_p^+ and σ_J are orthogonal to each other. BCP properties exhibit a similar pattern to that of the σ_p^+ values, producing relatively poor correlations for the first three sets but producing an excellent model for the set containing lithium. Even though the BCP properties perform much better than σ_p^+ at modelling the datasets exhibiting non-Hammett behaviour, the statistics are significantly inferior to those obtained using σ_J as the descriptor. From this, it is evident that BCP properties are unable to capture non-Hammett behaviour. Nevertheless, the BCP properties do produce an excellent correlation for the lithium dataset, which exhibits strong conventional Hammett behaviour. Thus this experiment, although unsuccessful at capturing non-Hammett effects, reinforces the conjecture that QTMS can act as a replacement for the conventional σ constants.

QTMS, although unsuitable for reproducing non-Hammett behaviour, can be used to model the σ constant in radical molecules. Previously only closed-shell species have been studied with QTMS. These findings mean that QTMS, in addition to being an ideal tool to study chemical reactions in a similar manner to σ, could be applied to the area of radical reactions,[47] particularly in the area of toxicology, where many compounds display a toxicity characterised by parameters such as σ^+.[48]

Toxicity of benzyl alcohols

In their paper on the toxicology of benzyl alcohols Kapur *et al.*[49] state that there is evidence that these compounds exhibit toxicity *via* a radical mechanism. To test that possibility they studied the toxicity of 13 *para*-substituted benzyl alcohols on rapidly dividing cancer cells. In their QSAR analysis they could find no evidence for an electronic effect but showed that the cellular toxicity was associated primarily with hydrophobicity. The activity data [log(1/*C*)] are listed in Table 6. Note that the observables are not Hammett constants but the original activity data [log (1/*C*))]. The point of this QTMS test case is to show that it fails when Hammett constants fail as well. In other words, both Hammett constants and QTMS descriptors are independently confronted with the original log(1/*C*) data. In such negative test cases it would be meaningless to use Hammett constants as observables.

Using log *P* alone as a descriptor, Kapur *et al.* obtained a relationship with $r^2 = 0.87$ and $q^2 = 0.80$ (although only 11 compounds were included, as the $C(CH_3)_3$ and NH_2 analogues were found to be outliers). The homolytic bond dissociation energy (BDE) is a direct estimate of the energy for abstraction of a hydrogen by a phenoxy radical. The BDE values of phenols have been shown[50] to correlate well with σ^+. However, addition of the σ^+ term did nothing to improve the correlation, nor did the inclusion of steric parameters. From this, the authors concluded that the benzyl alcohols exert their cytotoxicity *via* a polar narcosis mechanism that is sharply delineated by hydrophobicity alone.

In the QTMS analysis only the parent (*i.e.* non-radical) form of the benzyl alcohols provided the descriptors. The QTMS analysis failed to produce a model at all for this set of compounds. The first latent variable formed by SIMCA-P was found to be statistically insignificant (that is, its q^2 value was less than 0.097). The strong negative result adds weight to the conjecture that QTMS is only picking up electronic effects. Indeed, QTMS is unsuitable for modelling datasets where lipophilicity is the dominant factor in determining the activity.

Herbicidal activity of 5-chloro-2,3-dicyanopyrazines

Nakamura *et al.*[51] have reported the herbicidal activity of a set of 5-chloro-2,3-dicyanopyazines against barnyard grass. This grass is in competition with rice crops and selectivity toxic agents are actively sought. Here the activity is expressed as the pI_{50} value, which is the negative logarithm of the molar extracellular concentration required for 50% growth inhibition. The common skeleton for the dataset is given in Fig. 11. Table 7 shows the

Table 7 Substituent and activity data for the dicyanopyrazines

Molecule	R	pI_{50}	Molecule	R	pI_{50}
1	H	−0.59	12	OPh	−0.07
2	Me	−0.03	13	SMe	−0.02
3	Et	0.06	14	SEt	0.01
4	*n*-Pr	−0.01	15	SPh	−0.6
5	*n*-Bu	−0.13	16	NHMe	−0.22
6	CH_2Ph	−0.2	17	NHEt	−0.01
7	Ph	−0.1	18	NH-n-Bu	0.5
8	OMe	−0.44	19	NMe_2	0.24
9	OEt	−0.1	20	NHPh	0.11
10	O–*n*-Pr	0.24	21	Cl	−0.53
11	O–*iso*-Pr	0.3			

Table 6 Substituent and activity data for the *para*-benzyl alcohols

Molecule	Substituent	$\log[1/IC_{50}]$[a]	Molecule	Substituent	$\log[1/IC_{50}]$
1	CN	2.46	8	$COOCH_3$	2.47
2	C_6H_5	3.48	9	H	2.34
3	$C(CH_3)_3$	2.97	10	CH_3	2.77
4	Cl	2.75	11	Br	3.22
5	SCH_3	2.82	12	OC_4H_9	3.52
6	NO_2	2.56	13	OCH_3	2.53
7	NH_2	2.82			

[a] The activity is defined as the negative logarithm of the molar concentration required to produce 50% inhibition of cell growth in 48 hours (IC_{50}).

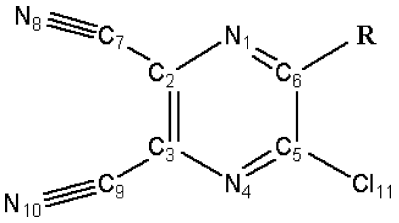

Fig. 11 Common skeleton and numbering scheme for the dicyanopyrazines.

substituent and activity data of the 21 compounds. Nakamura *et al.* examined this set using π^2 and π as lipophilic descriptors and σ_p as an electronic descriptor, obtaining a model with an r^2 of 0.891. Hence one might conclude that electronic effects are of significance for this dataset. Hansch, however, has also reformulated[34] this QSAR, and suspected that the σ_p term was "at least in part, a correction on π". His judgement was based on the fact that lipophilic parameters used in Nakamura's study were derived from the benzene system and these parameters are not readily transferable to the pyrazine system (occurring in the title compound). If this is the case, then the electronic term may not be of physical significance and the activity is wholly dependent on the lipophilicity. QTMS can provide an independent opinion in order to resolve this quandary.

A QTMS analysis,[21] without GA, yields a validated model of 1 LV with $r^2 = 0.33$ and $q^2 = 0.15$. A second analysis, now with a GA, results in a 1-LV model with $r^2 = 0.46$ and $q^2 = 0.35$, selecting for inclusion in the PLS analysis the Laplacian of the C_7N_8 and C_4N_5 bonds, and the ellipticities of the C_5Cl_{11} and C_9N_{10} bonds. Although a better model, even the GA cannot lift the predictivity of the model above random (since q^2 is still smaller than 0.5). It is clear that BCP properties fail to serve as meaningful descriptors from which a highly predictive model can be constructed. Based on previous test cases BCP properties emerge as purely electronic descriptors, and hence QTMS supports Hansch's view that only lipophilic descriptors are in charge.

At this point a further question[21] suggests itself. If σ_p acts merely as a correction term, why do the BCP variables themselves not act in the same manner? The answer lies in the default selection criteria for the LVs employed in SIMCA–P. A PLS regression that includes the BCP variables and the π and π^2 terms produces only one LV. If we subject Nakamura's original data (π, π^2 and σ_p parameters) to PLS analysis we find that only two LVs are produced by SIMCA-P's selection criteria and an r^2 of 0.47 is found. A third LV is needed to give the QSAR reported by Nakamura. However, the third LV is not significant within our cut-off limits. Hence the QSAR equation of Nakamura *et al.* is not validated by our criteria. This reason, combined with the likely physical insignificance of σ_p, explains why we are unable to obtain a good QSAR for the herbicidal activity of this series of molecules.

Chromatographic capacity factors of chalcones

Reversed-phase liquid chromatography has found application in QSAR studies as a rapid means of evaluating physicochemical properties of organic molecules, in particular log *P*. These properties are obtained from the chromatographic capacity ratios (denoted by k'_ϕ, where ϕ is the volume fraction of organic modifier in the eluent) by correlating the two sets of values *via* linear regression (eqn 2):

$$\log P = \mathrm{a} \log k'_\phi + b \qquad (2)$$

If the capacity factors are known then one can also obtain a value for the octanol–water partition coefficient. Luco *et al.*[52] have investigated the reversed-phase liquid chromatographic hydrophobicity parameters for several chalcone molecules in an effort to determine qualitative hydrogen bonding information from the chromatographic information obtained. The common skeleton for the chalcones is given in Fig. 12 and the substituent and activity data are given in Table 8. Here the activity QTMS attempts to model is log K'_w, the logarithm of the capacity factor extrapolated to 100% water in the eluent.

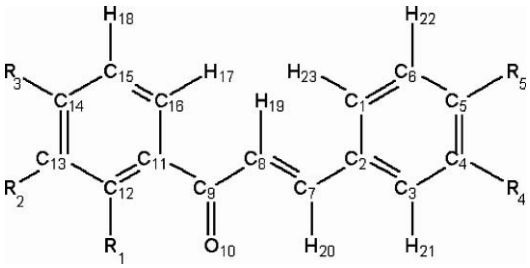

Fig. 12 Common skeleton and numbering scheme for the chalcones.

The combination of variables, included in the PLS analysis after GA variable selection, failed to produce any kind of model. Indeed, the first LV that SIMCA-P generated yielded $q^2 < 0.097$. Repeating the PLS analysis using all the BCP properties also failed to produce a model. Given the obvious importance of the hydrophobicity in determining the value of the dependent variable for this dataset, it is unsurprising that QTMS fails. Again, this result can be interpreted as support for the discriminatory power of QTMS and its inability to model hydrophobic effects.

Finally, it is worth mentioning that alternative quantum molecular similarity measures[53–57] are able to describe both polar substituents effects[58] and hydrophobicity.[59]

Conclusion

Many complex QSARs, whether medicinal or ecological, benefit from a multitude of Hammett constants to describe electronic effects. Although many Hammett constants have been listed, there is no guarantee that every possible substituent has been covered in every set of congeneric molecules. Moreover, since the Hammett constants are rooted in physical organic chemistry every application requires the use of the correct type

Table 8 Substituent and activity data for the chalcones

Molecule	R_1	R_2	R_3	R_4	R_5	log K'_w	Molecule	R_1	R_2	R_3	R_4	R_5	log K'_w
1	H	H	H	H	H	3.591	9	OH	H	H	OH	H	3.371
2	OH	H	H	H	H	4.147	10	OH	H	H	H	OH	3.371
3	H	H	H	H	OH	3.175	11	OH	H	H	H	OMe	4.169
4	H	H	H	H	F	3.639	12	OH	H	OMe	H	H	4.414
5	H	H	H	H	Me	4.072	13	OH	OMe	OH	H	H	3.664
6	H	H	H	H	NO_2	3.621	14	OH	H	H	H	Cl	4.646
7	H	H	H	H	Cl	4.139	15	OH	H	H	H	F	4.131
8	OH	H	OH	H	H	3.478							

of constant. This decision is typically guided by back-of-an-envelope arguments of traditional synthetic organic chemistry. With the advent of current computing power it is now possible to construct excellent QSARs overnight, linked to QCT descriptors directly drawn from modern solutions of the Schrödinger equation.

Early indications that QCT descriptors are highly correlated to Hammett constants were obtained from a set of carboxylic acids. In the current work, we have succeeded in proving that this assertion is more universal. We have shown that QTMS is capable of successfully modelling datasets where through-resonance effects are of importance. The range of Hammett parameters reproduced by QTMS has been extended to incorporate the σ^+ and σ^- constants, in addition to extending the range of molecular systems examined, including radicals. Meaningful active centers were highlighted. The evidence accumulated here suggests that QCT descriptors can most likely provide a reliable model to predict any type of Hammett constant.

Secondly, we have shown that QCT descriptors are purely electronic in nature. They fail to produce statistically valid QSAR models when lipophilicity/hydrophobicity are the ruling descriptors. The analysis of three sets (toxicity of benzyl alcohols, chromatographic capacity factors of chalcones, herbicidal activity of 5-chloro-2,3-dicyanopyrazines) show the discriminatory capacity of QCT descriptors. If they do not work then electronic effects are not in charge.

So, overall, when confronted with a measured activity (possibly of complex biological nature) the QCT descriptors have the correct flexibility to construct a reliable QSAR to predict this activity, and one is guaranteed that this is not because of steric effects of lipophilicity. A valid model can be obtained without invoking Hammett constants, thereby taking away the concern about which type of constant to invoke or whether the given substituent's constant is in the database.

Acknowledgements

Many thanks to an EPSRC CASE award sponsored by (then) SmithKlineBeecham and to Dr S. E. O'Brien for his preliminary work on the set of 5-chloro-2,3-dicyanopyrazines.

References

1 C. Hansch, A. Leo and R. W. Taft, *Chem. Rev.*, 1991, **91**, 165.
2 C. Hansch, W. E. Steinmetz, A. J. Leo, S. B. Mekapati, A. Kurup and D. Hoekman, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 120.
3 P. L. A. Popelier, *J. Phys. Chem. A*, 1999, **103**, 2883.
4 R. F. W. Bader, *Atoms in Molecules. A Quantum Theory*, Oxford University Press, 1990.
5 R. F. W. Bader, in *Encyclopaedia of Computational Chemistry*, ed. P. v. R. Schleyer, Wiley, Chichester, UK, 1998, 1, pp. 64–86.
6 R. F. W. Bader, *Chem. Rev.*, 1991, **91**, 893.
7 P. L. A. Popelier, *Atoms in Molecules. An Introduction*, Pearson Education, Harlow, UK, 2000.
8 R. F. W. Bader, *Pure Appl. Chem.*, 1988, **60**, 145.
9 R. F. W. Bader, *Phys. Rev. B*, 1994, **49**, 13348.
10 R. F. W. Bader and P. L. A. Popelier, *Int. J. Quantum Chem.*, 1993, **45**, 189.
11 P. L. A. Popelier, F. M. Aicken, and S. E. O'Brien, in *Chemical Modelling: Applications and Theory, Vol. 1*, Royal Society of Chemistry Specialist Periodical Report, ed. A. Hinchliffe, Ch 3, p. 143, 2000.
12 P. L. A. Popelier and P. J. Smith, in *Chemical Modelling: Applications and Theory, Vol. 2*, Royal Society of Chemistry Specialist Periodical Report, ed. A. Hinchliffe, Ch. 8, p. 391, 2002.
13 S. E. O'Brien and P. L. A. Popelier, *J. Chem. Inf. Comput. Sc.*, 2001, **41**, 764.
14 P. L. A. Popelier, in *Quantum topological molecular similarity-past, present and future*, ed. M. Ford, D. J. Livingstone, J. Dearden, and H. van de Waterbeemd, Oxford, UK, 2003.
15 S. E. O'Brien and P. L. A. Popelier, *J. Chem. Soc., Perkin Trans. 2*, 2002, 478.
16 P. L. A. Popelier, U. A. Chaudry and P. J. Smith, *J. Comput. Aided Mol. Des.*, 2004, **18**, 709.
17 P. J. Smith and P. L. A. Popelier, *J. Comput. Aided Mol. Des.*, 2004, **18**, 135.
18 U. A. Chaudry and P. L. A. Popelier, *J. Phys. Chem. A*, 2003, **107**, 4578.
19 P. L. A. Popelier, U. A. Chaudry and P. J. Smith, *J. Chem. Soc., Perkin Trans. 2*, 2002, 1231.
20 S. E. O'Brien and P. L. A. Popelier, ECCOMAS Proceedings, Barcelona, Spain, 2000.
21 S. E. O'Brien, *Quantum Molecular Similarity, an Atoms in Molecules Approach*, PhD Thesis, Dept. of Chemistry, UMIST, Manchester, Great Britain, 2000.
22 U. A. Chaudry and P. L. A. Popelier, *J. Org. Chem.*, 2004, **69**, 233.
23 B. K. Alsberg, N. Marchand-Geneste and R. D. King, *Chemom. Intell. Lab. Syst.*, 2000, **54**, 75.
24 B. K. Alsberg, N. Marchand-Geneste and R. D. King, *Anal. Chim. Acta*, 2001, **446**, 3.
25 P. L. A. Popelier, *Chem. Phys. Lett.*, 1994, **228**, 160.
26 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, *GAUSSIAN 98 (Revision A.7)*, Gaussian, Inc., Pittsburgh, PA, 1998.
27 *MORPHY98*, a program written by P. L. A. Popelier with a contribution from R. G. A. Bone, UMIST, Manchester, England, EU, 1998.
28 *UMETRICS, in SIMCA-P 10.0*, http://www.umetrics.com, Umeå, Sweden, 2002.
29 S. Wold, M. Sjostrom, and L. Eriksson, in *Encyclopaedia of Computational Chemistry*, ed. P. Schleyer, Wiley, Chichester, UK, 1998, vol. 3, pp. 2006–2021.
30 D. J. Livingstone, *Data Analysis for Chemists*, Oxford University Press, 1995.
31 The program SIMCA-P contains a predefined criterion for determining the significance of a Latent Variable (LV). If the q^2 of a newly constructed LV is less than 0.097 then that variable is considered unimportant to the model and is discarded. No more LVs are computed once this occurs and the PLS regression is completed. In order to obtain a valid model it is sometimes necessary to gradually reduce the number of LVs until the randomisation test is passed. This is justified as there is no point in attempting to interpret a model, even it does possess high values of r^2 and q^2, if it has been obtained purely by chance.
32 *SPSS Inc., version 10.0.7*, Chicago, USA, http://spss.com, 2000.
33 B. M. Wise, N. B. Gallagher, R. Bro, and J. M. Shaver, in *PLS Toolbox, Eigenvector Research, Inc.*, USA, 2002.
34 C. Hansch and A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology (Computer Applications in Chemistry)*, American Chemical Society, Washington, USA, 1995.
35 L. Amat, E. Besalu, R. Carbo-Dorca and R. Ponec, *J. Chem. Inf. Comput. Sci*, 2001, **41**, 978.
36 P. G. Mezey, *Mol. Phys.*, 1999, **96**, 169.
37 H. C. Brown and Y. Okamoto, *J. Am. Chem. Soc*, 1958, **80**, 4979.
38 C. Walling, A. L. Rieger and D. D. Tanner, *J. Am. Chem. Soc*, 1963, **85**, 3129.
39 P. J. Smith, *Quantum Chemical Topological Properties as Electronic Descriptors in Quantitative Structure–Activity/Property Relationships*, PhD Thesis, Dept. of Chemistry, UMIST, Manchester, Great Britain, 2003.
40 H. Kubinyi, in *Encyclopaedia of Computational Chemistry*, ed. P. v. R. Schleyer, Wiley, Chichester, UK, 1998, vol. 1, pp. 448–459.
41 S. H. Unger and C. J. Hansch, *J. Med. Chem.*, 1973, **16**, 745.
42 J. March, *Advanced Organic Chemistry, 4th edn*, Wiley, Chichester, UK, 1992.
43 L. Liu, Y. H. Cheng, Y. Fu, R. Chen and Q. X. Guo, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 1164.
44 M. Szwarc, C. H. Leigh and A. H. Sehon, *J. Chem. Phys.*, 1951, **19**, 657.
45 R. I. Walter, *J. Am. Chem. Soc.*, 1966, **88**, 1923.
46 R. A. Jackson, *J. Organomet. Chem*, 1992, **437**, 77.

47 C. Hansch and H. Gao, *Chem. Rev.*, 1997, **97**, 2995.
48 C. D. Selassie, R. Garg, S. Kapur, A. Kurup, R. P. Verma, S. B. Mekapati and C. Hansch, *Chem. Rev.*, 2002, **102**, 2585.
49 S. Kapur, A. Shusterman, R. P. Verma, C. Hansch and C. D. Selassie, *Chemosphere*, 2000, **41**, 1643.
50 P. Mulder, O. W. Saastad and D. Griller, *J. Am. Chem. Soc*, 1988, **110**, 4090.
51 A. Nakamura, O. Ikeda, H. Segawa, Y. Takeuchi and T. Takematsu, *Agric. Biol. Chem.*, 1983, **47**, 2923.
52 J. M. Luco, L. J. Yamin and H. F. Ferretti, *J. Pharm. Sci.*, 1995, **84**, 903.
53 R. Ponec, L. Amat and R. Carbo-Dorca, *J. Comput. Aided Mol. Des.*, 1999, **13**, 259.
54 R. Ponec, L. Amat and R. Carbo-Dorca, *J. Phys. Org. Chem.*, 1999, **12**, 447.
55 R. Carbo, E. Besalu, L. Amat and X. Fradera, *J. Math. Chem.*, 1995, **18**, 237.
56 L. Amat, R. Carbo-Dorca, D. L. Cooper, N. L. Allan and R. Ponec, *Mol. Phys.*, 2003, **101**, 3159.
57 L. Amat, R. Carbo-Dorca and R. Ponec, *J. Comput. Chem.*, 1998, **19**, 1575.
58 R. Carbo-Dorca, D. Robert, L. Amat, X. Girones, and E. Besalu, *Molecular Similarity in QSAR and Drug Design*, Springer, Berlin, Germany, 2000.
59 L. Amat, D. Robert, E. Besalu and R. Carbo-Dorca, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 624.