European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS 2000
Barcelona, 11-14 September 2000

QUANTUM MOLECULAR SIMILARITY: USE OF ATOMS IN MOLECULES DERIVED QUANTITIES AS QSAR VARIABLES

Sean E. O'Brien and Paul L. A. Popelier[*]

Dept. of Chemistry, U.M.I.S.T., 88 Sackville Street, Manchester M60 1QD,
England. e-mail: pla@umist.ac.uk, Web Page: http://www.ch.umist.ac.uk/morphy

Key Words: BCP Space, Atoms in Molecules, Quantum Molecular Similarity, QSAR, PLS

Abstract. We demonstrate how quantum mechanically derived data can be used in QSAR. Quantum Molecular Similarity Measures have utilised the total electron density with which to compare molecules. We have been able to use the properties of Bond Critical Points (BCPs) as defined by the topological theory of Atoms in Molecules (AIM) to accurately summarise and characterise any molecular electron density in a compact manner. Expanding earlier ideas we use 9 properties of the electron density, as measured at BCPs, as QSAR variables. These variables are determined for a series of congeners for which some physical property or activity has been measured. Using the Partial Least Squares (PLS) methodology the variables are fitted to this activity and a QSAR with predictive power is produced.

Additionally, PLS highlights the variables which are most important in predicting the activity. To simplify interpretation of the QSAR, we reduce each variable set which is centred on the same BCP to its first Principal Component (PC). These provide us with a quantum chemical fingerprint of a region of molecular electron density which can be directly linked to the chemist's intuitive view of a bond.

We apply this methodology to various groups of molecules and cover a range of properties. Physical properties are predicted and interpreted as well as biological and environmentally relevant toxicities and biodegradabilities.

1 INTRODUCTION

Quantum Molecular Similarity measures[1] have traditionally examined molecular electron densities in order to quantify how similar one molecule is to another. The ground state electron density is an obvious area in which to search for similarities, as Hohenberg and Kohn established that it underlies all molecular properties[2]. The central idea in this field being that similar molecules exhibit similar properties[3]. By assessing how alike two or more molecules are, we hope to be able to predict their properties and even to predict the properties of, as yet, unsynthesised molecules. This would obviously be of use in medicinal drug design and the fine chemical and agrochemical industry.

Conventional Quantative Structure-Activity Relationships (QSAR)[4] have sought to explain and predict activities of series of congeners by utilising empirical descriptors. These data reflect the molecular structure can be obtained from observation of structure, e.g. connectivity tables, experimental measurement, e.g. lipophilicity and substituent constants or semi-empirical calculations, e.g. HOMO-LUMO energies and molecular volumes[3]. These are mathematically fitted to maximally explain the property or activity of interest. There are numerous different descriptors which are and have been used for QSAR work.

The calculation of molecular electron densities by *ab initio* methods opens the prospect of calculating specific molecular properties for species that either do not exist, or have properties which are problematic to measure. The totality of the information contained within the density makes it an ideal candidate for comparing molecules. There have been notable successes when comparing complete electron densities[5-8] but there have been some problems associated with their use. The magnitude of the densities at the nuclear maxima are many times in excess of those in the chemically more interesting bonding zones. The similarity measures obtained can be severely biased towards the nuclei. Studies involving only valence electrons[9] and momentum densities[10] have sought to address this problem. Another area of difficulty is the superimposition of two densities. There are many methods available to do this[11-13] but they are computationally expensive and can lead to multiple maxima. Overall the comparison of total molecular electron densities is highly computer intensive and although many variations on this theme have been proposed[14-16] no one method has established itself as being obviously superior.

In previous publications[17-19] we have proposed the application of the theory of Atoms in Molecules (AIM)[20] to the field of Quantum Molecular Similarity. We show how properties derived from the topological analysis of the electron density can be combined with techniques used in traditional, empirical QSAR work. The application of rigorous *ab initio* data to structure-activity studies can provide a prediction of molecular properties.

2 BOND CRITICAL POINT PROPERTIES

The theory of AIM provides us with a definition of Bond Critical Points (BCP)[20,21]. These are saddle points in the electron density (ρ) and appear between any two nuclei which are said to be chemically bonded. They are stationary point in the gradient vector field of the density ($\nabla\rho = \mathbf{0}$) and appear at a minimum in ρ in the direction parallel to the internuclear axis and are at a maximum in all directions perpendicular to this axis. Figure 1. demonstrates the appearance of BCPs in phenol.

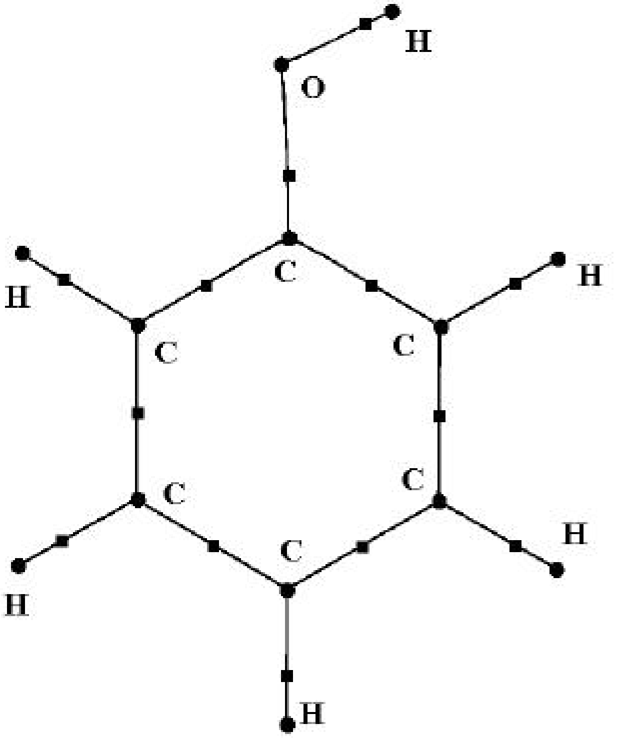


Figure 1. BCPs (represented as squares) appearing between all bonded nuclei in phenol.

Using the program MORPHY98[22] we are able to sample the electron density at each BCP and evaluate certain properties of ρ. These properties have been extensively studied[18,20,21] and have been seen to accurately and compactly characterise the properties of their associated bonds. We currently examine nine different properties at each BCP which are: ρ_b - the electron density, $\nabla^2\rho_b$ - the Laplacian of the electron density, λ_1,λ_2, λ_3 the three curvatures of the density (three eigenvalues of the Hessian of ρ), ε_b - the ellipticity (defined as λ_1/λ_2-1 and used as a shape descriptor), K_b and G_b - two kinetic energy densities and R_e -the equilibrium internuclear distance. In terms of the properties at a BCP, bondlength can be thought of as the sum of the distances to each of its linked nuclei. More details about each of the various properties can be found in refs. 20 and 21.

Any bond can be represented in terms of its BCP properties. Figure 2. shows the phenol molecule as represented in a two dimensional BCP space.

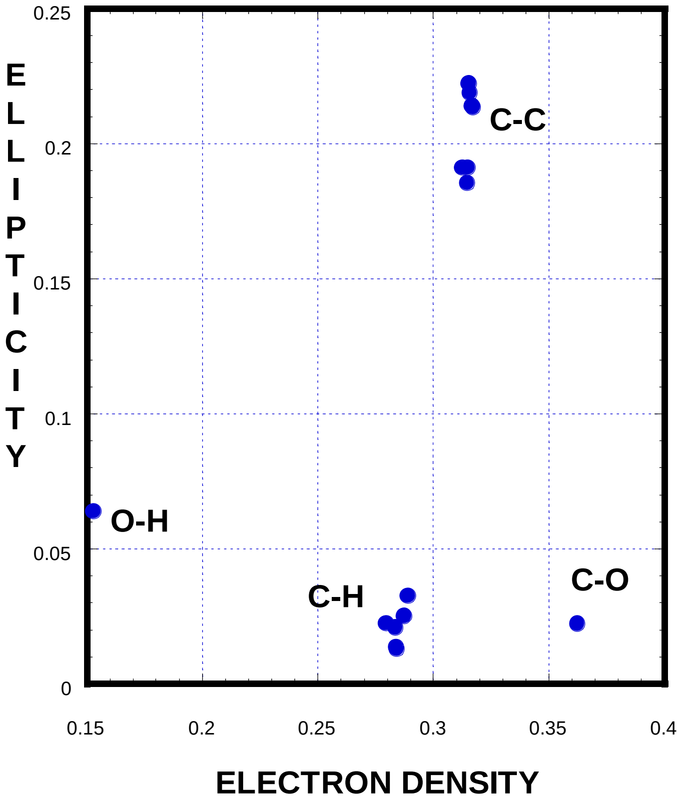


Figure 2. A representation of phenol when appearing in a two dimensional BCP space. Ellipticity is being plotted versus ρ at each BCP.

Any combination of BCP properties could have been used in the figure above. The BCPs all appear in distinct regions of the hyperspace. Similar types of bonds are in similar areas of BCP space. Phenol is seen as it appears in BCP space. The figure above only shows a two dimensional space but we actually use nine properties to represent any molecule in a nine dimensional hyperspace. It is by comparing molecules as they appear in BCP space that we proceed to formulate QSARs.

The use of BCP properties for molecular similarity work has some obvious advantages over the use of the full density. The process of obtaining the density from *ab initio* calculations is the same in both instances, but the subsequent analysis is quite different. The location of BCPs and the evaluation of their properties is computationally inexpensive. There is no problem associated with molecular alignment. The obstacles presented in maximising the overlap of molecular electron densities do not appear in BCP space. As the Bond Critical Points occur at the minimum in ρ along the internuclear axis, the density in this region is not overshadowed by the nuclear densities. They appear in the areas of the molecule at which chemical activity is more interesting. This is not to say that BCP properties are not affected by the core densities. Indeed, in many semi-empirical calculations where the nuclear densities are not explicitly modelled, BCPs do not appear. Their properties depend on the accuracy of the whole molecular electron density.

By representing any molecule as it appears in BCP space we circumvent several problems associated with Quantum Molecular Similarity work. We utilise BCP properties in the same manner as traditional QSAR descriptor variables to predict molecular properties and activities.

3 METHOD

The electronic wavefunction of the molecules of interest are calculated by *ab initio* methods. The only restriction on the choice of series is that they must share a common molecular skeleton. This is due to the large number of variables which would be generated if no bond to bond matching was decided *a priori*. At present this number is computationally unfeasible to deal with. However, there is no theoretical reason why such matching could not be done. Analysis of the electron densities produces 9 properties at each BCP. Each of these is then used as a descriptor variable. Thus the whole molecular set is described by 9n variables; where n is the number of bonds present in the common molecular skeleton. These variables are regressed against the property of interest using the method of Partial Least Squares Projections to Latent Structures (PLS)[23-25].

PLS is ideally suited for this form of multivariate analysis as it is designed to cope with data which has many noisy and collinear variables[26]. An equation is found in which the activity is modelled by a linear combination of these variables. In this way we are able to predict the property for any similar molecule. The quality of the fitted equation can be judged from the correlation coefficient - r^2 and the cross-validated correlation coefficient - q^2. Randomisation of the response variables is also carried out to assess the likelihood of the correlation occurring by chance. If good correlations are obtained with incorrect data, the initial regression may be due to chance factors alone. Assessment of the QSAR's validity is essential before any results can be interpreted in any physical sense[26].

PLS highlights those variables which are able to 'explain' the property of interest. We do not have to choose where in the molecule to look or which properties to look with. Hence *a priori* knowledge of a reaction centre for any given activity is not necessary[27].

To simplify the interpretation of the QSAR, we group all the variables that describe the same region of space together[28]. Reduction of these 9 descriptors to a single variable is carried out by means of Principal Components Analysis (PCA)[29]. The representation of the molecules in BCP space is now done by means of variables which are the first PC of BCP space taken at each BCP. In the same way as before the PLS highlights the important variables for each activity or property examined. These can now be associated directly with a region of electron density which a chemist will recognise as a bond.

We recommend that all the descriptors available are used to carry out the regression for predictive purposes. A reduced set of variables arising from PCA of localised BCP properties should then be used to interpret more fully the PLS analysis. Hence the use of BCP space to compactly represent the electronic structure of any molecule enables us to carry out QSAR studies which recover the property of interest. The variables identified in the PLS analysis can be regarded as intuitively reasonable and easily interpretable to the chemist.

4 CALCULATIONS

All *ab initio* calculations have been carried out using Gaussian 98[30], running under a Sun Microsystems UNIX operating system The BCP analysis has been executed by MORPHY98[22] and the PLS analysis was performed by the SIMCA-P package[31], both running under Microsoft Windows NT. All *ab initio* calculations were carried out at the following levels of theory: HF/3-21G(d)//HF/3-21G(d), HF/6-31G(d)//HF/6-31G(d), B3LYP/6-311+G(2d,p)//HF/6-31G(d) and B3LYP/6-311+G(2d,p)// B3LYP/6-311+G(2d,p). As the trends in BCP properties are preserved between basis sets[18] and the results of the PLS analysis are essentially basis set independent[32] we only report those carried out at the highest level of theory.

5 RESULTS and DISCUSSION

We now apply the method to various example sets.

5.1 Substituent constants for *para* and *meta* substituted benzoic acids

A series of 15 *para* and *meta* substituted benzoic acids was prepared and an investigation into their structure versus their relative acidity was carried out.

Table i. Substituents and Hammet substituent constants for substituted benzoic acids.

substituent	s para	s meta
NMe2	-0.83	-0.16
NHCH3	-0.7	-0.21
NH2	-0.66	-0.16
OCH3	-0.27	0.12
CH3	-0.17	-0.07
CH2CH3	-0.15	-0.07
CHCH2	-0.04	0.06
H	0	0
F	0.06	0.34
SH	0.15	0.25
Br	0.23	0.39
Cl	0.23	0.37
CF3	0.54	0.43
CN	0.66	0.56
NO2	0.78	0.71

Precisely because this set has been studied so often[3,33], they were found to be an ideal testing ground for our QSAR descriptors. The 15 molecules shown in Table i were fitted to their Hammet σ constants. All the molecules were computed in C_S symmetry with the exception of aminobenzoic acid in which the amino hydrogens were 20 degrees out of the plane.

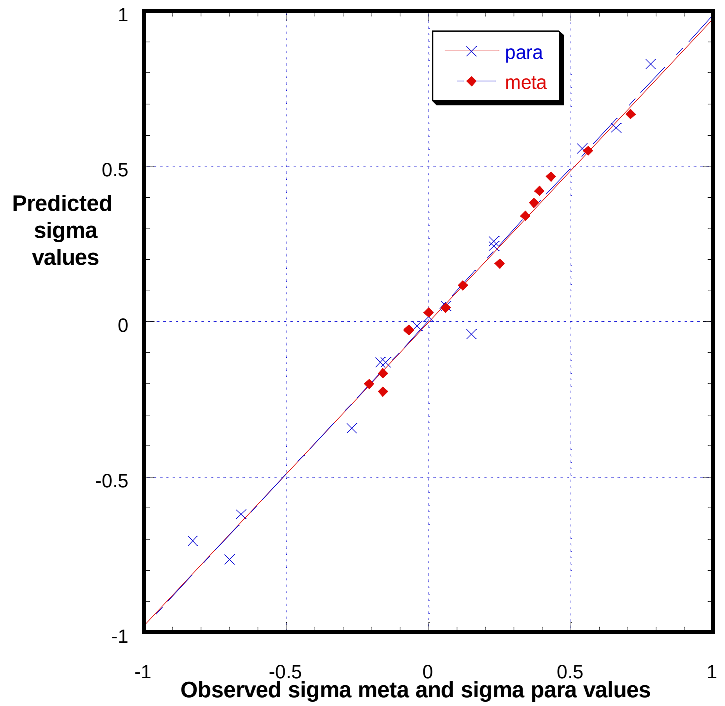


Figure 3. Observed versus predicted Hammett σ values for both *meta* and *para* substituted benzoic acids- using all the variables..

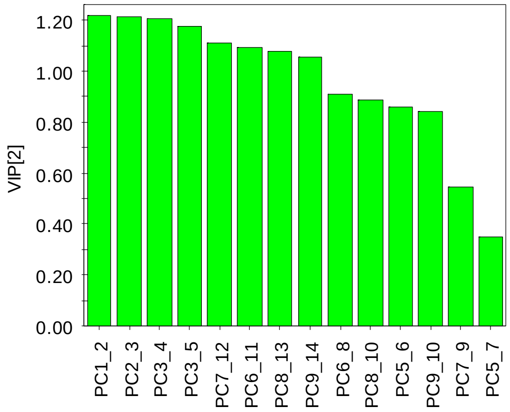
Using all the variables, the BCP variables were fitted to both the Hammet σ_p and σ_m constants. Table ii. shows the excellent correlations that are obtained, and also the internal predictive power as shown by q^2 for both substituent sets. Figure 3 illustrates graphically the correspondence between observed and predicted σ values.

Table ii. Correlations found when using PLS to form a QSAR, with different variable sets

variables used	r^2	q^2
s_p **-all variables**	0.98	0.96
s_p **-first PC**	0.97	0.95
s_m **-all variables**	0.99	0.94
s_m **- first PC**	0.99	0.98

Having recovered the electronic substituent effects within the molecular electron density, we examine the QSAR results. Taking the first PC of the 9 variables at each BCP location, we repeat the PLS analysis. The regression results are similar to those achieved with all the variables, demonstrating that we have not lost any information in extracting PCs. Figure 4 is a

Variable Importance Plot (VIP) for the 14 PCs used in the *para*-substituted benzoic acid QSAR. The height of each column demonstrates the relative importance of the particular variable in describing the property of interest.



Variable Importance Plot for para-substituted benzoic acids in a QSAR versus Hammett sigma values.

Simca-P 8.0 by Umetrics AB 2000-03-16 10:09

Figure 4. Variable Importance Plot

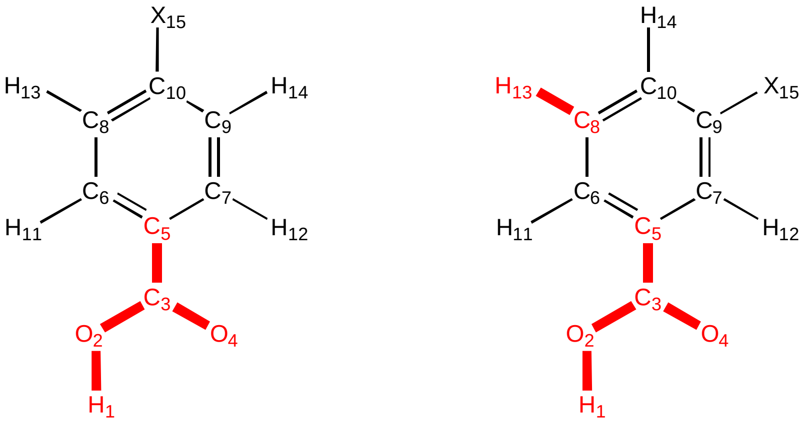


Figure 5. Numbering scheme for the substituted benzoic acids. X marking the site of substitution. The bonds marked in red are those deemed most important in the VIP (Figure 4). Those in blue are of secondary importance in describing the activity.

When viewing figures 4 and 5 one can see how the relative importance of each variable as determined by the PLS analysis is translated into a chemically relevant form. Figure 5 shows the regions of the molecule which are responsible for the observed variations in activity. In this case PLS picks out the carboxylic group as being chemically important for acidity. This is in agreement with all that we have learned about the reaction by using traditional methods. The *meta*-substituted case also picks out the *meta* C-H bond as having a large influence upon the acidity. The reasoning for this is as yet unclear and merits further investigation.

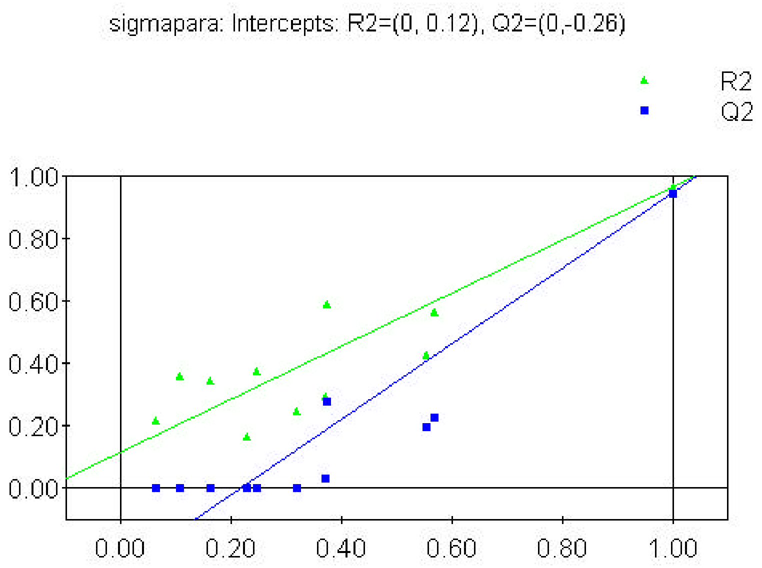


Figure 6. Plot of r^2 and q^2 obtained following randomisation of the response variable (σ). The Y axis is the value of the correlation coefficient and the X is the correlation of the original response variable with its permutation. Each point arises from a new permutation of the Y variable. Results for *para*-substituted benzoic acids.

Validation of any QSAR is essential if there is to be any meaningful interpretation of its results[26,29]. Figure 6 shows the values of r^2 and q^2 which arise if the same PLS analysis is carried out, but the response variable is randomised[26]. Hence each time the analysis is carried out (ten times in this case) each molecule is randomly assigned a value from the pool of Y variables. Any good correlations obtained would then be due to random factors or chance. If the values of the correlation coefficients recovered due to the permutation of Y variables are consistently below those obtained by the real analysis, one can be certain of the validity of the PLS results. It is clear from figure 6 that any conclusion made from the QSAR for *para*-substituted benzoic acids is founded on genuine correlations with the properties of the electron density. Similar validations were also noted for the *meta*-substituted benzoic acids.

From the examples above we can see that the use of BCP properties as QSAR variables in PLS analysis provides us with more than just predictive power. For substituted benzoic acids we also obtain an indication of the regions of the molecule which affect the activity. There is no need to pick out any particular zone of the molecule for examination. The PLS analysis is designed to do just that.

5.2 ^{13}C NMR chemical shifts in *para*- and *meta*-substituted benzonitriles

Carbon-13 NMR shifts have been reported for a series of substituted benzonitrile compounds[34].In order to explore whether BCP space could recover NMR data we formulated a QSAR with the chemical shift found for the C≡N carbon. The set of eight *para* and *meta* substituted compounds shown in Table iii. were prepared with the molecular skeletons depicted in Figure 7. All these results come from the HF/3-21G(d) level of theory.

Table iii. Substituents and 13C NMR shifts for substituted benzonitriles

Substituent	para-CN shift	meta-CN shift
H	118.81	118.81
CH_3	119.09	118.98
CH_2Cl	118.37	118.28
F	118.08	117.53
Br	118.01	117.26
Cl	117.93	117.40
OH	119.64	118.65
CF_3	117.46	117.36

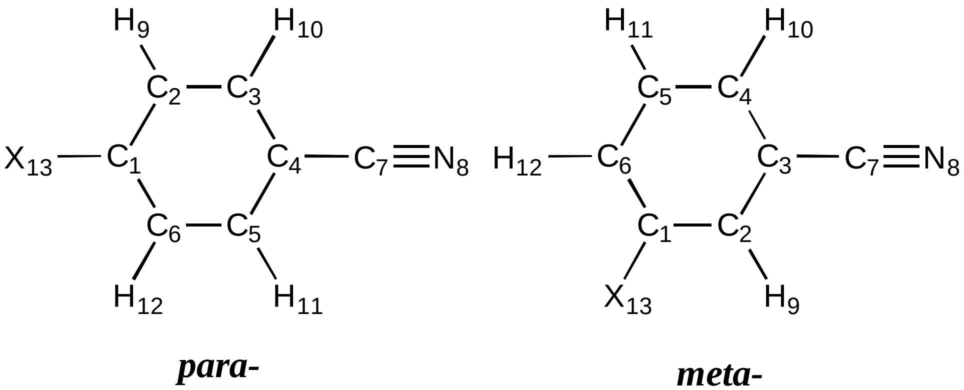


Figure 7. Molecular skeleton for the benzonitriles.

The correlations between experiment and the BCP variables are very high as can be seen in table 4 and the graph in figure 8.

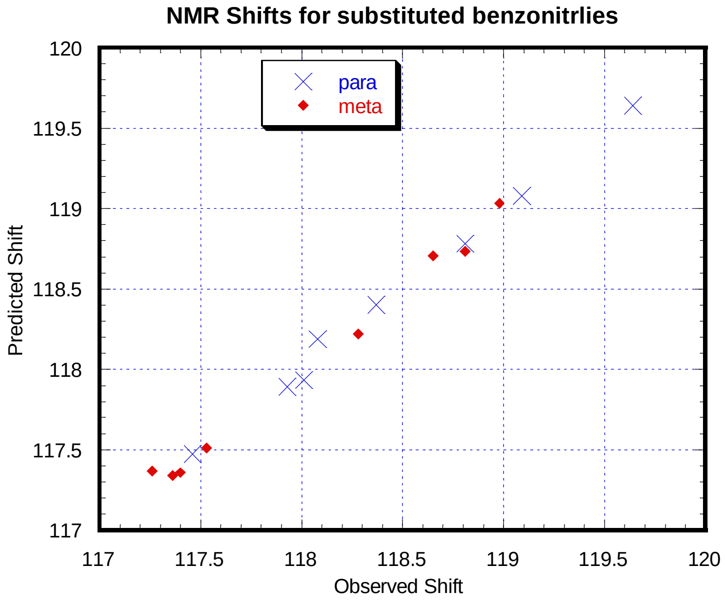


Figure 8.

Table iv. Correlations of BCP variables with ^{13}C NMR shifts of the cyano Carbon

Variables	r^2	q^2	bonds
All -para	0.99	0.77	N/A
1st PC -para	0.99	0.87	4-7 7-8
All -meta	0.98	0.87	N/A
1st PC -meta	0.98	0.88	3-7 1-13 7-8

5.3 Various properties of *para*-substituted phenols

A series of *para*-substituted phenols has been studied by Damborsky and co-workers[35]. They were assessed for biodegradability and toxicity. Among the descriptors used were pKa, molecular weight and the logarithm of the octanol/water partitioning coefficient $\log K_{ow}$, as seen in Table v. We tested the ability of the BCP variables to reproduce each of these properties.

Table v. Substituents and properties of the substituted phenols

Substituents	$\log K_{ow}$	Mw (Da.)	pKa	log 1/IGC50 (mol/L)	$\log K_b$ (L.org^{-1}.h^{-1})
H	1.48	89.07	9.92	-0.241	-11.16
Br	2.63	173.01	9.45	0.5	-11.8
CH_3	2.12	108.14	10.1	-0.162	-11.33
Cl	2.48	128.56	9.38	0.402	-11.77
CN	1.6	119.12	7.96	0.516	-13.82
NO_2	1.85	139.11	7.15	1.42	-13
OCH	1.57	124.14	10.2	-0.143	-12.7
$COCH_3$	1.45	136.15	8.05	-0.093	-12.51

The lipophilicity , molecular weight and acidity information was obtained from the literature[36,37]. Details of the toxicity and biodegradability data are available in ref 35. The values for toxicity are the 50% growth inhibitory concentration(IGC_{50}) and the biodegradability is the second order rate constant of oxidation of compound (K_b). As is customary in linear free-energy relationship studies, we examine the logarithms of these quantities.

Figure 9 shows the molecular skeleton for the molecular series and table vi gives the correlations obtained for all the QSARs. We show that BCP space reliably reproduces four of the five properties. The lack of any reasonable fit with molecular weight is included to demonstrate that we do not always achieve good correlations. Indeed one would not expect to recover the trends in molecular weight from the sampling of points exclusively within the common molecular skeleton.

Table vi. Correlations obtained with BCP variables when fitted to the properties below.

Variables	$\log K_{ow}$	Mw (Da.)	pKa	log 1/IGC50 (mol/L)	$\log K_b$ (L.org^{-1}.h^{-1})
r^2 - All variables	0.98	0.40	0.98	0.94	.99
q^2 - All variables	0.89	0.06	0.95	0.81	.90
r^2 - 1st PC	0.96	0.39	0.98	0.92	0.98
q^2 - 1st PC	0.82	0.07	0.95	0.79	0.84

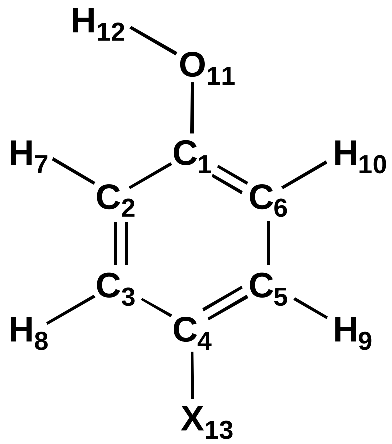
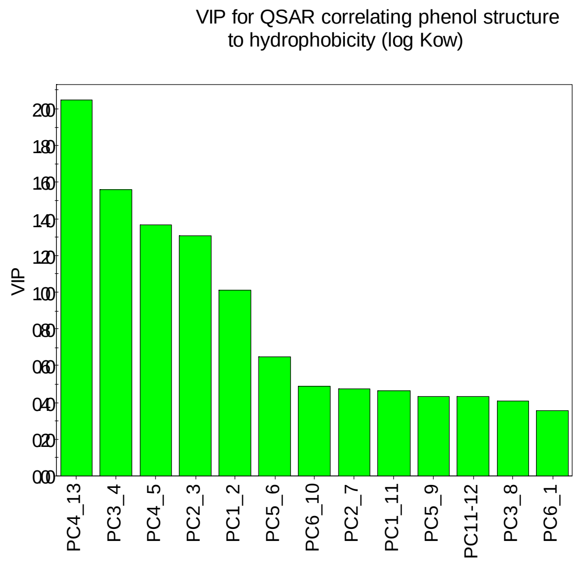


Figure 9. Molecular skeleton of the *para*-substituted phenol molecules X- denotes the site of substitution

Figures 10-13 are the Variable Importance Plots for each respective QSAR. They show that different regions of the molecular electron density are being highlighted each time. Hence without any *a priori* knowledge of the systems or mechanisms of action or propagation we are able to reproduce the properties of interest. In fact it is remarkable that these four properties which are not closely related (as can be seen in the correlation matrix table vii) can all be modelled by the same group of descriptors, which all arose from only 13 selected points in the electron densities.

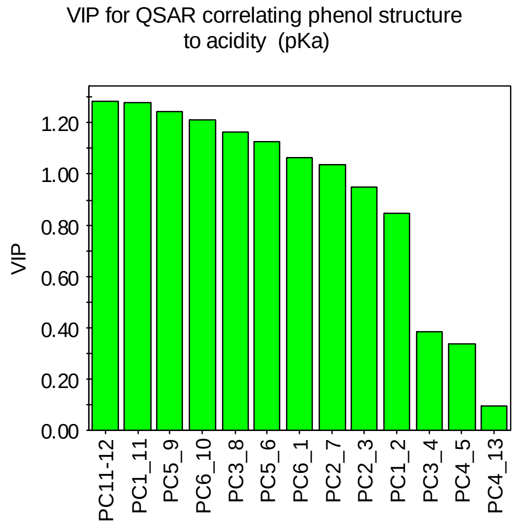
Table vii. Correlation matrix of properties modelled - values as r^2.

	logK_{ow}	**pKa**	**log 1/IGC50**	**logK_b**
logK_{ow}				
pKa	0.05			
log 1/IGC50	0.09	0.54		
logK_b	0.15	0.494	0.24	



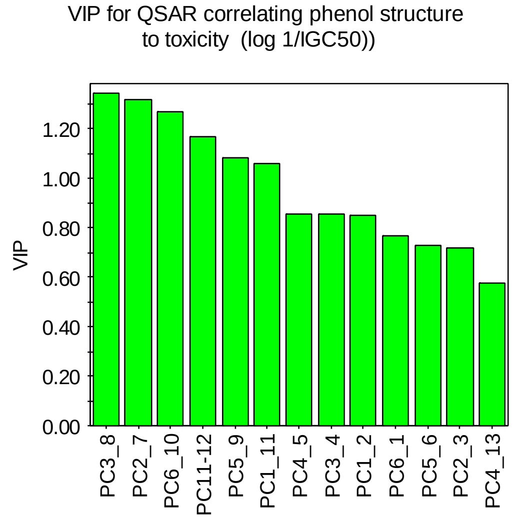
Simca-P 8.0 by Umetrics AB 2000-03-24 16:31

Figure 10.



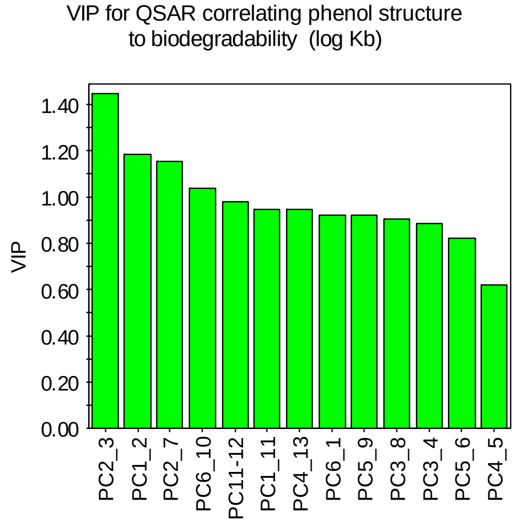
Simca-P 8.0 by Umetrics AB 2000-03-24 16:35

Figure 11.



Simca-P 8.0 by Umetrics AB 2000-03-24 17:29

Figure 12.



Simca-P 8.0 by Umetrics AB 2000-03-24 16:43

Figure 13.

6 CONCLUSIONS

We have shown how QSAR variables can be easily and efficiently extracted from the electron density. By combining AIM derived quantities and the PLS methodology we can accurately reproduce molecular properties and activities. The method is both computationally inexpensive and conceptually simple. There is a minimum of information which needs to be provided *a priori*. Only the common molecular framework has to be identified.

BCP space compactly summarises the molecular electron density. It provides a unique representation of each species and provides a method of comparison which is orientation independent. There are no problems associated with molecular superimposition. BCPs are located away from the huge nuclear densities and in the regions of chemical interest. In addition, the reduction of BCP space, by the extraction of principal components, provides a simple link to the chemists' idea of a bond.

PLS highlights the variables that are of most importance in any particular QSAR. For each property of interest PLS emphasises the regions that have the greatest significance. As was seen for the substituted phenols, the variables 'picked out' change when a different activity is being tested. The reasons behind the highlighting of particular areas merits further investigation.

Any property or activity of a group of similar compounds that can be attributed to the electronic structure can be modelled without prior knowledge of the mode of action. As such the activities of novel molecules and groups of molecules can be reproduced and insight obtained from examination of the molecular electron density.

REFERENCES

1)Carbo, R.; Leyda, L.; Arnau, M. *International journal of quantum chemistry* **1980**, *17*, 1185.

2)Hohenberg, P.; Kohn, W. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140*, 1133-1138.

3)Hansch, C.; Leo, A. Exploring QSAR : Fundamentals and Applications in Chemistry and Biology; 1 ed.; ACS: Washington, 1995; Vol. 1.

4)Kubinyi, H. *Quantitative Structure-Activity Relationships in Drug Design*, Ed., 1998.

5)Ponec, R. Similarity Models in the Theory of Pericyclic Macromolecules. *Topics in Current Chemistry* **1995**, *174*, 1-26.

6)Ponec, R. Similarity approach to chemical reactivity. Spin recoupling in chemical reactions. *Croatica Chemica Acta* **1998**, *71*, 501-510.

7)Rouvray, D. H. Similarity in Chemistry **-** Past, Present and Future. *Topics in Current Chemistry* **1995**, *173*, 1-30.

8)Amat, L.; Carbo-Dorca, R.; Ponec, R. Molecular quantum similarity measures as an alternative to log p values in QSAR studies. *Journal of Computational Chemistry* **1998**, *19*, 1575-1583.

9)Bowen-Jenkins, P. E.; Richards, W. G. Molecular Similarity in Terms of Valence Electron-Density. *Journal of the Chemical Society-Chemical Communications* **1986**, 133-135.

10)Allan, N. L.; Cooper, D. L. Momentum-Space Electron-Densities and Quantum Molecular Similarity. *Topics in Current Chemistry* **1995**, *173*, 85-111.

11)Nissink, J. W. M.; Verdonk, M. L.; Kroon, J.; Mietzner, T.; Klebe, G. Superposition of molecules: Electron density fitting by application of Fourier transforms. *Journal of Computational Chemistry* **1997**, *18*, 638-645.

12)Nelder, J. A.; Mead, R. simplex method. *Comput. J.* **1965**, *7*, 308.

13)Parretti, M. F.; Kroemer, R. T.; Rothman, J. H.; Richards, W. G. Alignment of molecules by the Monte Carlo optimization of molecular similarity indices. *Journal of Computational Chemistry* **1997**, *18*, 1344-1353.

14)Bowen-Jenkins, P. E.; Richards, W. G. Quantitative Measures of Similarity Between Pharmacologically Active Compounds. *International Journal of Quantum Chemistry* **1986**, *30*, 763-768.

15)Hodgkin, E. E.; Richards, W. G. Molecular Similarity Based On Electrostatic Potential and Electric- Field. *International Journal of Quantum Chemistry* **1987**, 105-110.

16)Lee, C.; Smithline, S. An Approach to Molecular Similarity Using Density-Functional Theory. *Journal of Physical Chemistry* **1994**, *98*, 1135-1138.

17)Popelier, P. L. A. , *Molecular Similarity in Drug Design*, Ed. P.M.Dean; Chapman and Hall: London, 1995, pp 213.

18)O'Brien, S. E.; Popelier, P. L. A. Quantum molecular similarity. Part 2: The relation between properties in BCP space and bond length. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **1999**, *77*, 28-36.

19)Popelier, P. L. A. Quantum molecular similarity. 1. BCP space. *Journal of Physical Chemistry A* **1999**, *103*, 2883-2890.

20)Bader, R. F. W. Atoms in Molecules: A Quantum Theory; Clarendon Press: Oxford, 1990.

21)Popelier, P. L. A. Atoms in Molecules: an introduction; Pearson: London, 1999.

22)Popelier, P. L. A. *Morphy98*; Popelier, P. L. A., Ed.: MORPHY98 - a program written by P.L.A. Popelier with a contribution from R.G.A. Bone. UMIST, Manchester, England, 1998.

23)Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. The Collinearity Problem in Linear-Regression - the Partial Least- Squares (Pls) Approach to Generalized Inverses. *Siam Journal On Scientific and Statistical Computing* **1984**, *5*, 735-743.

24)Wold, S. A Statistical View of Some Chemometrics Regression Tools - Discussion - Pls in Chemical Practice. *Technometrics* **1993**, *35*, 136-139.

25)Martens, H.; Naes, T. Multivariate Calibration; John Wiley & Sons: Guilford, 1989.

26)Wold, S.; Sjotsrom, M. *Partial Least Squares Projections to Latent Structures (PLS) in Chemistry*, Ed.; John Wiley & Sons: New York, 1998, pp 2006 - 2021.

27)Ponec, R.; Amat, L.; Carbo-Dorca, R. Molecular basis of quantitative structure-properties relationships (QSPR): A quantum similarity approach. *Journal of Computer-Aided Molecular Design* **1999**, *13*, 259-270.

28)Wold, S.; Kettaneh, N.; Tjessem, K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *Journal of Chemometrics* **1996**, *10*, 463-482.

29)Livingstone, L. Data Analysis for Chemists; first ed.; Oxford University Press: oxford, 1995.

30)Gaussian98; Revision A7. *M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J.Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker , J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.* **1998**.

31)UMETRICS *SIMCA-P 8.0*; UMETRICS, Ed.; info@umetrics.com: www.umetrics.com, 1998.

32)O'Brien, S. E.; Popelier, P. L. A. Quantum molecular similarity. Part 4: Incorporation of BCP Space into a QSAR framework. *Journal of Computational Chemistry* **2000**, *Submitted*.

33)Hammett, L. P. Physical Organic Chemistry; 2nd ed.; McGraw-Hill: New York, 1970.

34)Exner, O.; Budesinsky, M. Correlation of C-13 Substituent-Induced Chemical-Shifts Revisited **-** Meta-Substituted and Para-Substituted Benzonitriles. *Magnetic Resonance in Chemistry* **1989**, *27*, 27-36.

35)Damborsky, J.; Schultz, T. W. Comparison of the QSAR models for toxicity and biodegradability of anilines and phenols. *Chemosphere* **1997**, *34*, 429-446.

36)Perrin, D. D.; Dempsey, B.; Serjean, E. P. pKa Prediction for Organic Acids and Bases; Chapman and Hall: London, 1981.

37)Dean, J. A. Lange's Handbook of Chemistry; 14th ed.; McGraw-Hill: New York, 1992.