Quantum topological molecular similarity. Part 4.† A QSAR study of cell growth inhibitory properties of substituted (*E*)-1-phenylbut-1-en-3-ones

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

Dept. of Chemistry, U.M.I.S.T., 88 Sackville Street, Manchester, UK M60 1QD.
E-mail: pla@umist.ac.uk; Fax: +44-161-2004559; Tel: +44-161-2004511

Received (in Cambridge, UK) 5th November 2001, Accepted 17th December 2001
First published as an Advance Article on the web 24th January 2002

In this paper we apply a novel method called 'quantum topological molecular similarity' (QTMS) to a QSAR of antitumor activity of fifteen (*E*)-1-phenylbut-1-en-3-ones. The electronic structure of the molecules is compactly and accurately described by a set of topological descriptors drawn from *ab initio* wave functions. These descriptors consist of quantum mechanical properties evaluated at so-called bond critical points (BCP). These are special saddle points in the electron density located inside the molecule, roughly between two bonded nuclei. We use a partial least squares (PLS) analysis to obtain a valid regression with $r^2 = 0.91$ and $q^2 = 0.86$. QTMS highlights a region in the molecule comprising the active center of a Michael addition that has been surmised to be responsible for the mode of activity. This hypothesis is now independently confirmed.

Introduction

The ability to predict and explain drug activity has long been a goal of the pharmaceutical industry. Of the multitude of approaches available,[1] quantitative structure–activity relationships (QSARs) have been a useful and successful tool.[2–4] Traditionally a variety of 2-dimensional, empirical parameters have been used to equate structure with activity (Hansch analysis[2]). More recently, approaches that incorporate 3-dimensional data (CoMFA[5]) have proved popular. The use of small sets of parameters may provide insight into modes of action, whereas larger sets of parameters enable the best possible correlations to be found. However, at this moment the latter approach provides results that can be difficult to interpret in any mechanistic, physical sense.[6] Therefore there is an ongoing search for methods that enable mechanistic insight in addition to the prediction of activity.[2]

Over the last few years we have developed a novel 3D-QSAR method called quantum topological molecular similarity (QTMS),[7] whose action radius is currently being explored.[8,9] This method[10] aims to represent a molecule accurately and compactly based on the topology[11,12] of its electron density, denoted by ρ. The topology of ρ indicates special points in real 3D space, at which quantum chemical properties are evaluated, such as ρ itself, its Laplacian or a kinetic energy density. The representation of a set of molecules in the space of topological properties[13] can be correlated with measured properties, such as acidities of benzoic acids.[10] The idea of using quantum topology has been adopted by Alsberg *et al.*[14] in a study of the lowest UV transition for a system of 18 anthocynidins.

The use of the electron density as the ultimate source of information on a molecule is justified by the Hohenberg–Kohn theorem.[15] This theorem forms the basis of modern density functional theory (DFT)[16–18] and states that ρ is directly responsible for all of a molecule's ground state observable properties. Hence by comparing molecular electron densities the similarity between two or more molecules can be determined. This approach (and similar approaches[19–27]) has proved its merit[28–31] and is known as molecular quantum similarity (MQS).[32] The novel application of topological analysis to MQS, in the QTMS approach yields substantial savings in computer time and, potentially, an increased understanding of the mode of action. For a detailed discussion on the differences between QTMS and MQS in terms of advantages and limitations we refer to ref. 7

In this paper we apply QTMS to the antitumor activity of (*E*)-1-phenylbut-1-en-3-ones[33] and compare the results to those obtained with more traditional QSAR descriptors. We predict both the activity and the region of the molecule responsible for the observed activity. Hence, by quantifying the similarity between the aforementioned molecules, we fulfil the dual objectives of QSAR, both predicting and providing insight into biological activity.

The topology of the electron density

Over the last three decades a new theory[34] has been developed called 'atoms in molecules' (AIM),[11,12] which retrieves chemical insight from electronic wave functions. Although AIM has wide scope and application[35,36] we review only the AIM concepts that are of immediate relevance.

Bond critical points (BCP) are defined as points in real 3D space where the gradient of ρ vanishes and where the Hessian of ρ has two negative eigenvalues and one positive one. Thus, BCPs are saddle points in ρ, whose existence is dictated by the topology of ρ. They occur roughly in between nuclei that are conventionally regarded as bonded. They are mathematically defined points that are locatable in any high quality electron density (*ab initio* or X-ray derived).

We can characterize any molecule by a number of properties evaluated at its BCPs. The precise nature and function of these properties[37–39] is extensively discussed in ref. 7, but they can be associated with common chemical concepts such as bond strength, covalency, ionicity, shape and π-character.[11,12] In addition, the equilibrium bond length, designated by R_e is

† See ref. 7 for Part 3.
‡ *Current address*: Cancer Research Laboratories, University of Nottingham, Univ. Park, Nottingham, UK NG7 2RD.

 DOI: 10.1039/b110080g

added to the list of topological descriptors. The dependence of topological descriptors on R_e has been studied in detail before.[40] This study concluded that in general BCP properties cannot be trivially recovered or even predicted *via* R_e alone. Hence, R_e can in principle be viewed as an independent descriptor.

The QTMS analysis

At present we discern four stages in a QTMS analysis. The first stage is obtaining an electronic wave function for each molecule at a suitable geometry and level of theory.[7] This is the most time consuming step and the speed is dependent on the level of calculation employed. On a contemporary workstation (EV6 α-processor, 667 MHz, 1 Gb RAM) the generation of a typical wave function [3-(F_3C)C_6H_4] at B3LYP/6-311+G(2d,p)//HF-6-31G(d) level took about 5 hours of CPU time, of which 4 hours were required to compute the single point B3LYP wave function; the further part of the QTMS analysis taking less than ten minutes of computing time.

In the second stage the wave functions are passed on to (a local version of) the computer program MORPHY98,[41] which locates the BCPs using an automatic and robust algorithm.[42] In this step a 'discrete quantum fingerprint' is constructed, based on the topological descriptors.

In the third stage the descriptors are regressed against experimentally obtained activity data using the technique of partial least squares projections to latent structures (PLS),[43] as implemented in the program SIMCA-P.[44] We use the default criterion of SIMCA[45] to determine the optimum number of latent variables and the quality and validity of the regression are then assessed. These data may subsequently be used to predict the activity of related molecules.

The fourth stage focuses on the location of the active center, by running a new PLS analysis on a reduced set of variables. These variables are constructed *via* principal component analysis (PCA)[46] of the original BCP properties. By extracting PCs associated with each BCP, we summarize the electronic properties of that particular bond without discarding data. Only the PCs that have eigenvalues greater than one are extracted using the program SPSS.[47] The percentage variance explained by each PC at every BCP is typically around 90% or higher. It is important to realize that PLS is carried out again, this time on the extracted (principal components) rather than on the 'raw' variables (BCP properties). The variable importance in the projection (VIP) values show the relative importance of each independent variable ('*X*') in the regression. Therefore, factors that contribute considerably to the fit have high VIP scores. It is a working hypothesis of QTMS that the active center of a molecule consists of the BCPs associated with the highest VIPs. This idea arises as it is the properties of ρ, at these particular BCPs, which best explain the activity of interest. It may occur that the active center is rather diffuse, in that the VIP plot shows only a gradual diminishing of the scores. In addition, PLS may highlight as important variables that have no physical significance. However, in the current case these problems do not arise.

It would be wrong to object to the way we use PCs to locate the reactive center by pointing out that PCs by their nature encapsulate information about the entire system, explaining the variation of the entire data set. In fact, as explained in detail in ref. 7, a PCA is carried out for each BCP *separately* and yields one or more PCs for each BCP. Since each BCP is described by a set of local properties there is no doubt about the validity of identifying local PCs with BCPs.

Finally it should be mentioned that QTMS enables the inclusion of potentially important parameters that are not represented *via* BCP properties, such as the populations of topological atoms. Hence, the topological properties can be viewed as just one set of possible QTMS descriptors.

Data generation

Another research group has collected all the experimental data used in this work. The α,β-unsaturated ketone, (*E*)-1-(4′-hydroxyphenyl)but-1-en-3-one, was extracted[33] from the dried plant *Scutellaria barbata* D. Don (*Labiatae*), which has been used in traditional Chinese medicine as an antitumor agent. It was found to possess moderate antitumor activity and hence a series of substituted analogues was synthesized. The antitumor activity was determined by using an *in vitro* cell culture system (MTT assay) against the K562 human chronic myelogenous leukaemia cell line. The IC_{50} concentration was calculated with reference to a standard growth curve and represents the concentration required to cause a 50% decrease in cell growth after five days' incubation.[33] Seventeen substituted butenones and their IC_{50} values are listed in Table 1, comprising the present QSAR set. The most active compound (1.9 μM) has an activity of 1.7 orders of magnitude over the least active one (90 μM). The rest lie somewhere in between these two, with a slight bias toward more active molecules. Note that the measured IC_{50} values were taken from ref. 33 and the phenylbutenones' lipophilicity was calculated by the ClogP[48] program. The atom-labelling scheme is shown in Fig. 1.

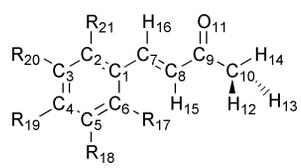

Fig. 1 Labeling scheme of the general molecular skeleton of the set of phenylbutenones. The aryl group consists of a phenyl ring with five possible substitution sites (R_{17-21}). The conformation of the terminal methyl is as it appears in the lowest energy geometry.

The molecular wave functions for the first 15 molecules shown in Table 1 were computed at the B3LYP/6-311+G(2d,p)//HF-6-31G(d) level of theory,[49,50] as implemented in the program GAUSSIAN98.[51] It has been shown[40] that this split level of calculation is adequate for reproducing BCP properties. We have previously extensively investigated the issue of basis set variation and it was shown that patterns between BCP properties are preserved between basis sets.[40] In general, the area of the active center is also conserved[7] in several benzoic acids derivatives. However, it was also seen that the order of the VIPs, the precise value of the correlation coefficients and the number of latent variables found may differ. In addition, we compare the results with those found using the semi-empirical AM1 Hamiltonian.[52] Due to the approximations used by the semi-empirical method in order to reduce computational effort, they do not always produce reliable topologies. Hence, when using AM1 derived wave functions we utilize only the equilibrium bond lengths. Good results have been seen with this level of calculation (at a fraction of the computational expense of *ab initio* methods) but very poor results have also been observed. It has been proposed that *ab initio* level calculations should be employed in conjunction with AM1 to ensure the accuracy of the method for a given data set. If AM1 provides consistently good results, further investigation can be done at lower cost and on larger variants of the system.[7] The semi-empirical calculations performed similarly to the *ab initio* levels for this system. Hence, we subsequently computed the properties of two further phenylbutenones [4-MeC_6H_4 and 2,4-$(MeO)_2C_6H_3$] at the AM1 level. These two compounds were not included in the initial analysis due to SCF convergence difficulties.

Geometry optimizations with various initial conformations were carried out. All geometries collapsed to the planar configuration, giving the molecule C_s symmetry. However, the conformation in which the ring plane is at 90° to the plane of the side chain proved to be a transition state (as confirmed by a

Table 1 Growth inhibitory activities (IC_{50} values/μM) and ClogP values for the set of seventeen substituted (*E*)-1-phenylbut-1-en-3-ones

Aryl group	IC_{50}	ClogP
C_6F_5	1.9	2.78
3-$NO_2C_6H_4$	2.6	3.07
4-$NO_2C_6H_4$	2.6	1.81
3-Br-4-FC_6H_3	2.9	2.93
3-$(F_3C)C_6H_4$	3.5	2.95
5-Br-2-FC_6H_3	5.3	3.07
4-BrC_6H_4	5.6	1.81
4-FC_6H_4	6.1	2.21
4-ClC_6H_4	8.3	2.78
4-(MeS)C_6H_4	15	2.62
C_6H_5	17	2.07
3-OH-4-(MeO)C_6H_3	42	1.25
4-$(OH)C_6H_4$ [a]	60	1.4
4-$(MeO)C_6H_4$	79	1.98
4-$(Me_2N)C_6H_4$	90	2.26
Test set		
4-MeC_6H_4	7.5	2.56
2,4-$(MeO)_2C_6H_3$	30	2.07

[a] Naturally occurring compound.

vibrational frequency analysis). As such, the conformation was more than 20 kJ mol^{-1} higher in energy, all ensuing calculations were started with C_s symmetry but no subsequent symmetry constraints were imposed. The orientation of the terminal methyl group was shown to contribute ±4.1 kJ mol^{-1} to the overall energy, with the conformation having hydrogen (H_{14}) eclipsed with the oxygen being the more stable (Fig. 1). Every substituent group was given all probable starting geometries and in each case the lowest energy conformations were chosen.

All topological descriptors for each bond were generated as described above. We also evaluated the QSARs produced with traditional σ parameters and total atomic charges. The former descriptors were taken as the preferred values listed in ref. 2 and the latter from a Mulliken population analysis[53] as listed in the GAUSSIAN98 output. The charges used were calculated at the B3LYP/6-311+G(2d,p)//HF/6-31G(d) level of theory and for the further predictive work at the AM1 level.

The descriptors described above were regressed against the activity data shown in Table 1 using the method of PLS. The significance of each regression was assessed using the correlation coefficient r^2 and the cross-validated correlation coefficient,[46] q^2. The latter coefficient is dependent on the PRESS score, calculated here by leaving out approximately one seventh of the data.[45] Additionally, we subjected our regressions to a randomization (or permutation) of the data. This test estimates the likelihood that a good fit is obtained purely by chance. The original *Y* data are randomly permuted to appear in a different order. The reordered *Y* variables are related to the unperturbed, independent *X* variables by refitting the model. New r^2 and q^2 values are obtained for every permutation. If, in each case, the model based on the randomized variables has considerably lower r^2 and q^2 values, one can feel confident about the validity of the original model.[46]

In the last stage of the QTMS analysis, the principal components that had eigenvalues greater than one were extracted from the set of BCP properties. A second PLS analysis was performed on these PCs and the importance of each component was assessed.

As a final comment it should be pointed out that the present version of QTMS[7] does not, in principle, partition the data into a training set and a test set, although this option is currently being explored in our lab in connection with neural networks replacing PLS. A proper split of the data set into training and test set would require more data than we have access to in this paper. Nevertheless, we have treated two compounds (Table 1, entries 16 and 17) as a 'test set' although it should be made clear that the predictive power of our model is actually tested internally, *via* q^2, obtained after 'leaving out one seventh of the data'.

Table 2 Statistics obtained from the PLS regression of the 15 substituted (*E*)-1-phenylbut-1-en-3-ones' growth inhibitory activities [$\log(1/IC_{50})$] against the QTMS descriptors and Mulliken charges [at both B3LYP/6-311+G(2d,p)//HF/6-31G(d) and AM1 levels] and σ constants

	r^2	q^2	Valid	No. of cases
QTMS-B3LYP	0.91	0.86	YES	15
QTMS-AMl	0.89	0.85	YES	15
σ Constants	0.87	0.85	YES	13
Mulliken charges B3LYP	0.80	0.65	YES	15
Mulliken charges AM1	0.81	0.72	YES	15

Results and discussion

The results of the regressions of all the descriptors against the growth inhibitory activities of the 15 phenylbutenones are shown in Table 2. The QTMS analysis using the *ab initio* calculations displays the best fit but the distinction between the QTMS statistics and those found when using σ is marginal. Both sets of Mulliken charges are clearly less good at reproducing the activity data. All the regressions appear to be valid.

As two of the substituents appear *ortho* to the butenone chain, σ constants were not available for two of the molecules—including the most active (pentafluorinated) compound. This is a well-known drawback of using substituent constants.[46] Calculations can prove extremely instructive in those situations where empirical parameters are unavailable or difficult to determine experimentally.

An additional problem arises when utilising the Hammett σ parameter. In traditional QSAR, σ has proved to be very adept at explaining the electronic effects, and hence the different activities of sequences of molecules varying only in their substituents. However, problems can arise from the σ constants from one system being applied to all systems. In fact, this approximation only works for similar systems and so to combat this, a plethora of σ 'constants' have been obtained. The idea is that the type of σ measure that best fits the data (gives the best correlation) reveals information about the reaction mechanism. However, the sheer profusion of electronic parameters and the fact that substituent effects are not always additive are distinct drawbacks for this method. By using the predictive power of molecular orbital calculations we are able to effectively 'measure' the changes in molecular electronic properties. We are able to bypass parametrization altogether and avoid its inherent problems.

As can be seen from Table 2, the regression obtained with the AM1 semi-empirical calculations is similar to the *ab initio* result. This is not always the case, as mentioned above. However, when this does occur we have the opportunity to calculate further compounds at a vastly reduced computational cost. Hence, we performed predictions of the activities of two further compounds, the 4-MeC_6H_4 and 2,4-$(MeO)_2C_6H_3$ derivatives, whose antitumor activity had already been experimentally determined. The predictions were made using QTMS with AM1 calculations, Hammett σ parameters and AM1 derived Mulliken charges. The results of these calculations are shown in Table 3.

None of the IC_{50} predictions are absolutely in agreement with experiment, but in general they offer a reasonable estimate of activity. The obvious exception is the Mulliken charge based prediction of the 2,4-$(MeO)_2C_6H_3$ compound. As one of the substituents for this molecule is in the *ortho* position, Hammett σ's were also unable to predict the activity. It appears that

Table 3 Predictions of the activity of the 4-MeC_6H_4 and 2,4-$(MeO)_2C_6H_3$ derivatives, using QTMS-AM1, σ constants and AM1 Mulliken charges

	4-MeC_6H_4 $\log(1/IC_{50})$	$IC_{50}/\mu M$	2,4-$(MeO)_2C_6H_3$ $\log(1/IC_{50})$	$IC_{50}/\mu M$
Experimental	−0.88	7.5	−1.48	30.0
QTMS AM1	−1.31	20.6	−1.75	56.5
σ Constants	−1.40	24.9	Not available	Not available
Mulliken charges AM1	−1.22	16.4	−17.08	1.21×10^{17}

QTMS is able to evaluate the activity of this compound whereas the other two methods fail. Future work will encompass comparisons for other systems with more highly valued NBO charges.

It is highly desirable for any QSAR method to provide insight into the mode of action associated with a given activity. As previously described, QTMS extracts principal components from the BCP properties and uses these to locate the active centre of the molecules. Hammett σ parameters are able to reveal that the presence of electron withdrawing groups increases activity but are unable to indicate where in the molecule the activity might take place. By analysing the results of PLS we are able to find the variables that are most important for explaining the response data. The VIP plots obtained with the PCs and the Mulliken charges are shown in Figs. 2 and 3 respectively.

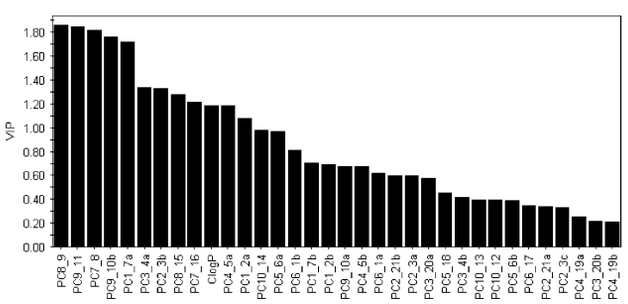

Fig. 2 Plot of the VIP scores of the PLS analysis on the principal components (PCs). The five highest scores are clearly separated from the others.

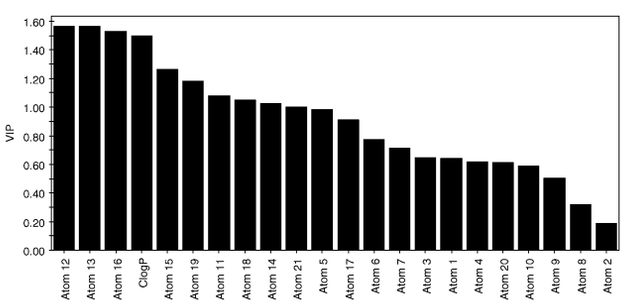

Fig. 3 Plot of the VIP scores of the PLS analysis on the B3LYP/6-311+G(2d,p)//HF/6-31G(d) Mulliken charges.

The most important variables for explaining the antitumor activity, when analysed by QTMS, are those shown in bold in Fig. 4. There is a clear division in the significance in the

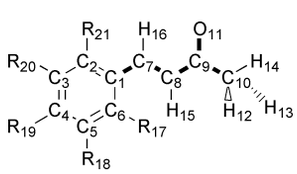

Fig. 4 The active center (bold) as found from the VIP plot (see Fig. 2).

regression between the first five variables and the rest (although such divisions are not always quite as clear[54]). The bonds highlighted encompass the length of the butenone chain. The variables picked out as being important by the Mulliken charge analysis are charge on the terminal hydrogen atoms, the charge on hydrogen 16 and the lipophilicity.

The mechanism of action of the series of butenones is unknown. However, it has been suggested that the mode of action 'may involve an alkylation process, possibly *via* a Michael addition'.[54] This conjecture is supported by the observation that the presence of electron withdrawing groups increases activity.[33] There are similar cases where Michael addition is thought to be responsible for cytotoxicity.[55] The mechanism of Michael addition is well known and indeed occurs in the very area that we recover as the active centre from the QTMS analysis.

The reaction scheme shown in Fig. 5 is that conventionally

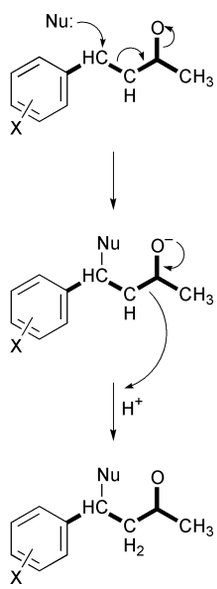

Fig. 5 Schematic representation of Michael addition involving a nucleophile and a phenylbutenone. The highlighted region is shown in bold.

associated with the Michael addition. The active center, as revealed by the PLS analysis is shown in bold. Encouragingly, the active center encompasses the expected bonds that undergo the reaction. Although the two peripheral bonds in bold, BCPs 1–7 and 9–10, are not directly involved in the mechanism the electron density in those regions is greatly affected by changes at carbons 7 and 9. It is not surprising, therefore, that they are highlighted. However, as can be seen in Fig. 2 they have a smaller VIP value than the bonds directly involved (BCPs 7–8, 8–9 and 9–11). Interestingly, the electron density in the CH bonds, BCPs 7–16 and 8–15, are not relevant to the explanation of activity, neither are the bonds in the terminal methyl group. We are not able to unequivocally confirm that cell growth inhibition is triggered by means of Michael addition, however the QSAR strongly concurs with the aforementioned hypothesis, based on experimental evidence.

Advantages and current limitations of QTMS

One of the difficulties of conventional QSAR arises from the need to have a data set, which consists of 'well behaved'

substituents. Parameters may not be known for unusual functional groups and of course σ constants are not necessarily transferable between systems. Certainly predicting the effects of *ortho* substituents can be problematic and it has been advised to stick to sets of well-characterized substituents.[2] This can lead to difficulties if the data set was not generated with QSAR in mind. The advantage provided by our approach is that no constraints are put on the make up of the data set. Any compound that can be computed *via* quantum chemistry (*ab initio* or even semi-empirically) can have its activity predicted (assuming the modes of action to be the same). This advantage has been illustrated before in the prediction of QTMS in the acidity of trisubstituted benzoic acids.[7]

Another benefit of QTMS is that it independently predicts the active center. The perennial problem has been which part of the density to examine.[29,56,57] Analysis of the total molecular density often hides the subtle changes responsible for small changes in activity and it is well known that reactions occur at specific sites within the molecule. The major differences found between total densities will more often than not arise within the differing substituents themselves. In an attempt to overcome this difficulty Ponec *et al.*[29] have also correlated electron densities of substituted benzoic acids with the Hammett σ constant by only looking at the COOH group. However, they state that this is only possible when 'the reaction centre can unambiguously be determined'. As the majority of QSAR work, especially in the field of drug design, deals with novel agents with unknown modes of action, the reaction centre will frequently be unknown. It is here that we 'bridge the gap' between conventional QSAR and quantum molecular similarity. Finally, within a certain class of applications there is no need for 3D superposition.

At the present stage of development QTMS experiences certain limitations. Currently QTMS has only been tested on predominately rigid systems showing marginal conformational flexibility. Other than requiring dramatically increased computing time the study of conformationally flexible molecules poses the question of how conformational change expresses itself in the topological descriptors. Although the 3D information due to conformational variations is implicit in BCP properties, the extent to which this affects QSARs is not yet fully understood. This area is now under investigation. Secondly, in its present formulation QTMS adopts a few caveats of QSAR, such as the necessity of sufficiently large sets of molecules and the fact their activity must be based on the same mode of operation. If QTMS is developed towards the more general area of molecular similarity the importance of these caveats may decrease. To date we have only investigated linear QSAR models and the extension of the method to incorporate non-linear models may well increase the range of applicability of the method. Finally, QTMS has only been applied to a set of molecules that share a common skeleton. Put more precisely this means that the mapping of BCPs between molecules is one-to-one and unambiguous. A relaxation of this prerequisite is subject to ongoing research.

Conclusion

Our aim in this paper is to demonstrate the effectiveness of QTMS in fulfilling the traditional aims of QSAR. This method combines the predictive power of modern *ab initio* calculations with QSAR *via* quantum topological descriptors and a rigorous statistical analysis of the regression. A compact and accurate description of the electronic structure of fifteen substituted (*E*)-1-phenylbut-1-en-3-ones was regressed against their measured antitumor capability. A PLS analysis showed that this electronic structure description correlates very strongly to the observed activity. Traditional σ constants also perform well but are not universally applicable. Analysis of Mulliken charges provides a QSAR with little predictability and no comprehensible insight into the mode of action. QTMS highlights a region in the molecules that strongly overlaps with the active center of the Michael addition. This fact corroborates a conjecture previously put forward based on experimental evidence.

Acknowledgements

We are grateful to Dr N. J. Lawrence, D. Rennison and Dr F. M. Aicken for useful discussions and to the referees for their constructive and helpful feedback. Appreciation is expressed to EPSRC for a studentship (GR/L65895).

References

1 P. M. Dean, *Molecular Similarity in Drug Design*, Chapman and Hall, London, 1995.
2 C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, 1st edn., ACS, Washington, Vol. 1.
3 H. Kubinyi, in *Quantitative Structure–Activity Relationships in Drug Design, Encyclopedia of Computational Chemistry*, Vol. 3, ed. P. v. R. Schleyer, Chichester, 1998, p. 2309.
4 C. Hansch and H. Gao, *Chem. Rev.*, 1997, **97**, 2995.
5 R. D. Cramer, S. A. DePriest, D. E. Patterson and P. Hecht, *3D QSAR in Drug Design*, ed. H. Kubinyi, ESCOM, Leiden, 1993.
6 C. Hansch, D. Hoekman and H. Gao, *Chem. Rev.*, 1996, **96**, 1045.
7 S. E. O'Brien and P. L. A. Popelier, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 764.
8 P. L. A. Popelier, P. Smith, U. Chaudry and S. E. O'Brien, 2001, unpublished results.
9 S. E. O'Brien, Quantum Molecular Similarity, An Atoms in Molecules Approach, PhD Thesis, UMIST, Manchester, 2000.
10 P. L. A. Popelier, *J. Phys. Chem. A*, 1999, **103**, 2883.
11 R. F. W. Bader, *Atoms in Molecules. A Quantum Theory*, Oxford University Press, 1990.
12 P. L. A. Popelier, *Atoms in Molecules. An Introduction*, Pearson Education, 2000.
13 P. L. A. Popelier, in *Molecular Similarity in Drug Design*, ed. P. M. Dean, Chapman & Hall, London, 1995, p. 215.
14 B. K. Alsberg, N. Marchand-Geneste and R. D. King, *Chemom. Intell. Lab. Syst.*, 2000, **54**, 75.
15 P. Hohenberg and W. Kohn, *Phys. Rev. B*, 1964, **136**, 864.
16 E. S. Kryachko and E. V. Ludena, *Energy Density Functional Theory of Many-Electron Systems*, Kluwer, Dordrecht, 1990.
17 R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, Oxford, 1989.
18 W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, Weinheim, 2000.
19 N. L. Allan and D. L. Cooper, *Top. Curr. Chem.*, 1995, **173**, 85.
20 P. J. Artymiuk, A. R. Poirrette, D. W. Rice and P. Willett, *Top. Curr. Chem.*, 1995, **174**, 73.
21 E. Besalu, R. Carbo, J. Mestres and M. Sola, *Top. Curr. Chem.*, 1995, **173**, 31.
22 P. G. Mezey, *Top. Curr. Chem.*, 1995, **173**, 63.
23 P. C. Mishra and A. Kumar, *Top. Curr. Chem.*, 1995, **174**, 27.
24 R. Ponec, *Top. Curr. Chem.*, 1995, **174**, 1.
25 D. H. Rouvray, *Top. Curr. Chem.*, 1995, **173**, 1.
26 Y. Takahashi, *Top. Curr. Chem.*, 1995, **174**, 105.
27 G. Tasi and I. Palinko, *Top. Curr. Chem.*, 1995, **174**, 45.
28 L. Amat, R. Carbó-Dorca and R. Ponec, *J. Med. Chem.*, 1999, **42**, 5169.
29 R. Ponec, L. Amat and R. Carbó-Dorca, *J. Comput.-Aided Mol. Des.*, 1999, **13**, 259.
30 L. Amat, R. Carbó-Dorca and R. Ponec, *J. Comput. Chem.*, 1998, **19**, 1575.
31 R. Ponec, L. Amat and R. Carbó-Dorca, *J. Phys. Org. Chem.*, 1999, **12**, 447.
32 R. Carbó-Dorca, D. Robert, L. Amat, X. Girones and E. Besalu, *Molecular Similarity in QSAR and Drug Design, Lecture Notes vol. 73*, Springer, Berlin, 2000.
33 S. Ducki, J. A. Hadfield, L. A. Hepworth, N. J. Lawrence, C. Y. Liu and A. T. McGown, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 3091.
34 R. F. W. Bader, P. L. A. Popelier and T. A. Keith, *Angew. Chem., Int. Ed. Engl.*, 1994, **33**, 620.
35 R. J. Gillespie and P. L. A. Popelier, *Chemical Bonding and Molecular Geometry from Lewis to Electron Densities*, Oxford University Press, New York, 2001.

36 P. L. A. Popelier, F. M. Aicken and S. E. O'Brien, in *Atoms in Molecules, Spec. Publ. R. Soc. Chem.*, ed. A. Hinchliffe, 2000, vol. 1, p. 143.
37 R. F. W. Bader, P. J. MacDougall and C. D. H. Lau, *J. Am. Chem. Soc.*, 1984, **106**, 1594.
38 R. F. W. Bader, T. S. Slee, D. Cremer and E. Kraka, *J. Am. Chem. Soc.*, 1983, **105**, 5061.
39 R. F. W. Bader and H. J. T. Preston, *Int. J. Quantum Chem.*, 1969, **3**, 327.
40 S. E. O'Brien and P. L. A. Popelier, *Can. J. Chem.*, 1999, **77**, 28.
41 P. L. A. Popelier, in MORPHY98, MORPHY98—a program written by P. L. A. Popelier with a contribution from R. G. A. Bone, UMIST, Manchester, England, 1998.
42 P. L. A. Popelier, *Chem. Phys. Lett.*, 1994, **228**, 160.
43 S. Wold, M. Sjostrom and L. Eriksson, in *Partial Least Squares Projections to Latent Structures (PLS) in Chemistry*, ed. P. von Ragué Schleyer, New York, 1998.
44 UMETRICS, in SIMCA-P 8.0, www.umetrics.com, 1998.
45 A. UMETRICS, SIMCA-P 8.0 User Guide and Tutorial, 1999.
46 L. Livingstone, *Data Analysis for Chemists*, Oxford University Press, 1995.
47 SPSS Inc., version 10.0.7 http://www.spss.com, SPSS Inc., 2000.
48 C. C. Biobyte Corp., in Biobyte Corp., Claremont California, CLogP Windows version 2.0, www.biobyte.com, 1995.
49 J. B. Foresman and A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian Inc., 1996.
50 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
51 GAUSSIAN98, Gaussian 98, Revision A.7, 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. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian, Inc., Pittsburgh PA, USA, 1998.
52 A. J. Holder, in *AM1, Encyclopedia of Computational Chemistry*, vol. 1, ed. P. v. R. Schleyer, 1998, p. 11.
53 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833.
54 N. J. Lawrence, A. T. McGown, S. Ducki and J. A. Hadfield, *Anti-Cancer Drug Des.*, 2000, **15**, 135.
55 A. C. Beekman, H. J. Woerdenbag, W. van Uden, N. Pras, A. W. T. Konings, H. V. Wikström and T. J. Schmidt, *J. Nat. Prod.*, 1997, **60**, 252.
56 C. Lee and S. Smithline, *J. Phys. Chem.*, 1994, **98**, 1135.
57 P. E. Bowen-Jenkins and W. G. Richards, *Int. J. Quantum Chem.*, 1986, **30**, 763.