Ester Hydrolysis Rate Constant Prediction from Quantum Topological Molecular Similarity Descriptors

U. A. Chaudry and P. L. A. Popelier*

Department of Chemistry, UMIST, Manchester, M60 1QD, England

Received: February 1, 2003; In Final Form: April 3, 2003

A previously established method [*J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764], called quantum topological molecular similarity, is applied to obtain an excellent and statistically validated quantitative structure−activity relationship (QSAR) of base-promoted hydrolysis rate constants for a set of 40 esters. This work is relevant for environmental exposure and risk analysis and proposes a reliable and cheaper alternative to measuring infrared group frequencies for that purpose. Our method draws descriptors from modern ab initio wave functions, which have become affordable by the current abundance of inexpensive computing power. We acquire a 3D geometry-optimized picture of each molecule and characterize its bonds further with four quantities defined by the theory of quantum chemical topology. Without molecular superposition we then construct a variety of models, which all point toward the molecular fragment (O═C)−C−O being most significant to explain the range of hydrolysis rate constants. This highlighted zone is called the active center, and when the model is confined to it, a QSAR of $r^2 = 0.930$ and $q^2 = 0.863$ is obtained for all 40 esters.

1. Introduction

Properties and activities of molecules in complex environments, ranging from pure solvents to active sites of enzymes, can often be related to their features in the gas phase in a highly quantitative way. The deeper reason behind this phenomenon is still obscure, but such quantitative structure−activity and structure−property relationships (QSARs/QSPRs) continue to be discovered by the thousands.[1−3] These gas-phase properties may be determined experimentally, such as infrared (IR) group frequencies, or they may be computed, such as HOMO−LUMO energy gaps. Regardless of the origin of the gas-phase descriptors, the outcome of their inevitable chemometric treatment strongly depends on the quality of the descriptors.

A plethora of theoretical descriptors of variable quantum chemical credibility have been proposed[4,5] and used in a wide variety of contexts. Sometimes these descriptors are defined ad hoc within an approximate theory, such as superdelocalizabilities[6] in the context of frontier orbital theory. Sometimes they are straightforward properties, such as molecular dipole moments but then often generated by crude semiempirical models (CNDO, MNDO, INDO) or even extended Hückel theory. With the current availability of very economical computer power one is in an ideal position to produce more realistic and less compromising quantum chemical descriptors. Modern ab initio methods, including the popular density functional theory (DFT),[7,8] can now be used to describe the electronic structure of gas-phase molecules of sufficiently large size to be practically important. In this work this approach is combined with a modern theory described as "quantum chemical topology" (QCT)[9,10] that enables the careful extraction of chemical insight from modern wave functions. In that capacity QCT is widely used[11,12] in areas from mineralogy to biochemistry to high-resolution crystallography.

There is an urgent need[13] for reliable and low-cost methods for predicting rate constants of organic compounds. Not only is this a prerequisite for environmental exposure and risk analysis,[14] but such rate constants (e.g., of hydrolysis) are also important to industries that try to improve speciality chemicals.[15] Esters are among the most common acid derivatives appearing in nature (e.g., fruity aromas) and medicinal (e.g., methyl salicylate) and industrial (e.g., adhesives, films, cleaners, polishes, and plastics) products.[16] Some time ago Collette proposed[17] a method for predicting reactivity parameters of organic chemicals from spectroscopic data to help assess the environmental fate of pollutants. In particular, he used midinfrared gas-phase spectra to predict the alkaline hydrolysis rate constants (k_{OH}) of 41 carboxylic esters, the subject of this paper. The environmental fate is known to be highly dependent on the polar and steric nature of the compound, which should hence be captured by Hammett σ constants and Taft steric parameters. The advantage of spectroscopically based property−reactivity correlations is that they can be determined in a rapid, inexpensive, and precise manner. In this work we go one step further by proving that there is not even a need to measure relevant IR frequencies. Instead we show that ab initio optimized geometries of the esters and their corresponding wave functions are sufficient to predict accurately their base-promoted hydrolysis.

The computational technique used for this purpose is called quantum topological molecular similarity (QTMS), which is explained in the next section. QTMS has delivered promising to superb QSARs before, such as the σ_p, σ_m, σ_I, and σ_p^0 parameters of mono- and polysubstituted benzoic acids, phenylacetic acids, and bicyclic carboxylic acids,[18] the toxicity and biodegradability of *para*-substituted phenols, ^{13}C NMR chemical shifts in *para*- and *meta*-substituted benzonitriles,[19] antibacterial activity of nitrofuran derivatives,[20] the binding affinity of the classical[21,22] steroid dataset for corticosteroid binding,[20] and many others. The current paper can be regarded as part 6 in our series on QTMS.

QTMS has also inspired closely related techniques such as StruQT,[23,24] while QCT has been used to predict hydrogen bond donor capacity[25] and hydrogen bond basicity.[26]

* To whom correspondence should be addressed. Fax: +44-161-2004559. Phone: +44-161-2004511. E-mail: pla@umist.ac.uk.

10.1021/jp034272a CCC: $25.00
Published on Web 05/10/2003

2. QTMS and Computational Details

Since a detailed account of the QTMS method can be found elsewhere,[27] we only review it here. In essence we first optimize a molecule's geometry by a spectrum of quantum chemical methods. If we chose to describe the molecule beyond its optimized bond lengths, we evaluate given functions (such as the electron density) in special points in the molecule. We then obtain a number of localized descriptors, which we confront with the experimental data (here k_{OH} rate constants) via a chemometric technique. QTMS does not involve any molecular superposition despite utilizing 3D information, a feature in common with comparative molecular moment analysis (CoMMA).[28] It shares with the benchmark[29] technique CoMFA[21] the ability to point out the "active site" of a ligand, namely, the place where local changes in ligand property (electrostatic potential, bond length, steric field) influence the compound's activity most. To some extent it shares with modern molecular similarity techniques[30] such as QMSM[31] the central role of the electron density.

On the basis of a practical guide[32] to the ab initio computer program we used[33] to optimize the molecular geometries, we selected five "levels of theory" labeled from A to E. This choice is established by an opportune property-accuracy/computing-cost ratio. In this paper we sustained the labeling for sake of consistency with previous[18,19,27,34−36] and future publications. Level A corresponds to the semiempirical model AM1,[37] which yields reasonable bond lengths for nonesoteric (hence, already parametrized) molecules but fails to provide an electron density that can be analyzed by QCT. All higher levels (B, C, and E) used in this work generate the "single-point" wave function at the optimized geometry. Level B is HF/ 3-21G(d),[32] while level C, still at the Hartree−Fock level, invokes a more complete basis set, or HF/6-31G(d). Omitting level D, we also obtained results at the most expensive level, E, B3LYP/6-311+G(2d,p), where electron correlation is modeled by a popular hybrid density functional.[38] In the current state of affairs ab initio work still needs to explore several levels of theory because of the inherent fluctuations in the data it provides, which is why we report in the next section the QSAR statistics for all four levels.

Now we discuss how a bond can be characterized beyond its equilibrium (or optimized) value R_e. According to QCT, there is a so-called bond critical point (BCP) for each bond. This is a special point in space, roughly between the two bonded nuclei, where the gradient of the electron density vanishes. It is the point of lowest density on the topological curve connecting the two bonded nuclei. Intuitively, the BCP can be seen as a valley between two mountain peaks where the peaks are the nuclei and the mountains are the (topological) atoms. As such the BCP acts as a boundary between two atoms and hence bears a signature of the bond that links them. Therefore, it makes sense to evaluate the electron density, denoted by ρ, at the BCP and use it as an extra descriptor next to R_e. The Laplacian of the electron density, or $\nabla^2\rho$, measures to what extent the electron density is locally depleted (when $\nabla^2\rho > 0$) or locally concentrated (when $\nabla^2\rho < 0$). When evaluated at a BCP, $\nabla^2\rho$ discriminates covalent bonds from ionic and van der Waals bonds, simply by its sign. If $\nabla^2\rho$ is negative, then one calls the bond covalent. Again the Laplacian can be included as an extra QCT descriptor. Another descriptor to join the list is the so-called ellipticity "ϵ", a rigorous definition of which is given elsewhere.[10] When zero, the ellipticity indicates that the bond is cylindrically symmetric. On the other hand, an increasingly nonzero value indicates ovality of the electron density contour lines, which is an expression of π-character in appropriate circumstances. A final QCT descriptor we considered is a type of local kinetic energy density, denoted by K. Of course, when computed for a Kohn−Sham-based density functional (e.g., B3LYP), K refers only to a noninteracting reference system. In summary, at level A, each bond is described by just R_e, and at the three other levels, B, C, and E, it is described by a vector containing four components (ρ, $\nabla^2\rho$, ϵ, K). We used a local version of the program MORPHY98[39] to obtain the latter QCT descriptors.

Models were constructed by using the partial least-squares (PLS)[40] method, as implemented in the program SIMCA-P.[41] This advanced multiple linear regression technique is designed to find a (linear) relationship between the observed Y variables (k_{OH}) and the $4N$ X variables or QCT descriptors, where N is the number of bonds included in the QTMS analysis. Note that for level A we only have N descriptors, namely, the optimized bond lengths themselves. The PLS technique has been designed to handle many X variables, potentially noisy and collinear. Although we do not involve as many X variables as a typical CoMFA analysis would, we still benefit from the latter two features of PLS. In general, the four descriptors ρ, $\nabla^2\rho$, ϵ, and K depend on R_e in a complex and nonlinear fashion.[34] To some extent, the small deviations the used approximate ab initio methods introduce in the values of the five descriptors can be interpreted as noise on the values achieved from exact wave functions.

The quality of the PLS regression is assessed by a number of criteria. First, classic measures[42] are the correlation coefficient r^2 and the cross-validated correlation coefficient q^2. The latter measures the internal consistency of the model by omitting[43] one-seventh of the data (Y variables) and predicting the missing data by a model based on the remaining Y variables. Second, we consider the number of latent variables (LVs) the PLS analysis constructs. SIMCA-P considers an LV to be significant if q^2 corresponding to a newly constructed LV is larger than 0.097. If q^2 drops under 0.097, the LV is discarded and the PLS analysis is completed. Although the debate has not been settled, we believe that a model with fewer LVs is superior to one with more. Finally, the model has to validate on the basis of a so-called randomization test. This test guards against "correlation by chance" by monitoring the deterioration of the model (measured by r^2 and q^2) as the Y variables are randomly permuted. SIMCA-P prescribes default cutoff values beyond which the model ceases to be valid.

We used the so-called variable importance in the projection (VIP)[43] to detect the active center of the compound. Descriptors (or X variables) with a VIP value smaller than 1 can be rejected as unimportant, whereas those with the highest VIP values constitute the active center. Although there are no rigorous statistical criteria to pinpoint the active center, we show in the next section that this course of action is very useful and makes sense in a practical context.

3. Results and Discussion

Table 1 shows the measured log k_{OH} values for the 40 esters we selected from Collette's original[17] 41 esters. Ethyl iodoacetate was not included since basis sets for iodine were not readily available. The rate constant we construct the QSAR for refers to the reaction

$$RC(=O)OR' + OH^- \rightarrow RC(=O)O^- + R'OH$$

where the polar nature of R strongly affects the receptivity of nucleophilic attack by OH^-.

TABLE 1: Alkaline Hydrolysis Rate Constants (k_{OH}, M^{-1} s^{-1}, 25 °C) of Ethyl-Containing (Left) and Other (Right) Esters

no.	compound	log k_{OH}[a]	no.	compound	log k_{OH}
1	ethyl *n*-butyrate	−1.26	17	methyl formate	1.56
2	ethyl isobutyrate	−1.49	18	benzyl acetate	−0.71
3	ethyl acetate	−0.96	19	*n*-butyl acetate	−1.06
4	ethyl benzoate	−1.50	20	*n*-propyl acetate	−1.06
5	ethyl bromoacetate	1.70	21	methyl acetate	−0.74
6	ethyl formate	1.41	22	isopropyl formate	1.04
7	ethyl chloroacetate	1.56	23	methyl benzoate	−1.10
8	ethyl acrylate	−1.11	24	methyl methacrylate	−1.25
9	ethyl 2-bromopropionate	1.00	25	benzyl benzoate	−2.10
10	ethyl *p*-fluorobenzoate	−1.41	26	isopropyl acetate	−1.52
11	ethyl dibromoacetate	2.31	27	*n*-butyl formate	1.34
12	ethyl *p*-nitrobenzoate	−0.13	28	*n*-propyl formate	1.36
13	ethyl *p*-aminobenzoate	−2.59	29	*sec*-butyl acetate	−1.76
14	ethyl trichloroacetate	3.41	30	2-chloroethyl acetate	−0.41
15	ethyl pivalate	−2.77	31	2-methoxyethyl acetate	−0.69
16	ethyl aminoacetate	−0.19	32	methyl *p*-fluorobenzoate	−1.15
			33	methyl *p*-hydroxybenzoate	−1.52
			34	methyl *p*-aminobenzoate	−2.35
			35	isopropyl hydroxybenzoate	−2.23
			36	methyl *m*-aminobenzoate	−1.47
			37	isopropyl *p*-aminobenzoate	−3.04
			38	methyl 2,4-D[b]	1.06
			39	2-butoxy 2,4-D[b]	1.48
			40	*n*-octyl 2,4-D[b]	0.57

[a] Original references containing these data can be found in ref 17. [b] 2,4-D is a well-known pesticide, also known as (2,4-dichlorophenoxy)acetate.

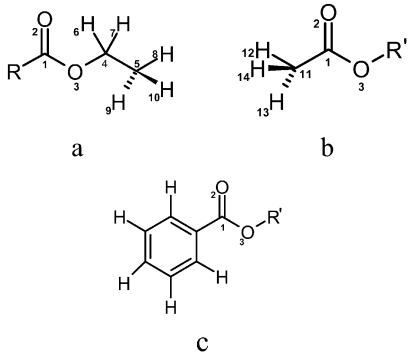

Figure 1. Numbering schemes for the esters: (a) the ethyl subset (R′ = CH_2CH_3), (b) the acetates (R = CH_3), (c) the benzoates (R = phenyl).

The applications of QTMS have hitherto taken advantage of the straightforward one-to-one mapping of bonds that a congeneric set of molecules typically offers. In other words, we have not yet studied sets of widely differing molecules but confined ourselves to molecules that have a common skeleton. For example, all esters have the (O═C)—O—C fragment in common, such that we can sample the QCT descriptors for that common skeleton from one molecule and map it onto the corresponding information in another molecule. In the PLS data matrix each row contains the *Y* variable (k_{OH}) and one or four QCT descriptors for each bond, for levels A and B and C and E, respectively. The order in which the descriptors appear is fixed and depends on the common skeleton one adopts for a given model.

Figure 1 shows the numbering scheme for the common skeleton of three possible subsets of esters. The largest subset of 16 esters is one where the alcohol fragment R′ coincides with the ethyl group (Figure 1a). The second largest subset contains 14 acetates where the carboxylic R group is methyl (Figure 1b). The third largest and final subset we considered encompasses 11 benzoates (R = phenyl, Figure 1c). There are many more possible subsets, such as the formates and the methyl-containing esters, but the small number of esters they contain would reduce the statistical quality of the ensuing model. The purpose of studying these subsets is to demonstrate that, overall, the active center settles for the three bonds of and around the central (O_2═C_1)—O_3—C_4 fragment, where the C_1—O_3 bond breaks upon hydrolysis. Then, in a second stage, we focus on the QCT descriptors in those three bonds with an eye on extracting chemically relevant information of the active center.

TABLE 2: Survey of the PLS Analyses Obtained for the Ethyl-Containing Esters

level	descriptor	LV	r^2	q^2
A	bond lengths	2	0.828	0.765
B	bond lengths	2	0.935	0.887
	BCP properties	2	0.930	0.897
C	bond lengths	2	0.969	0.948
	BCP properties	2	0.952	0.917
E	bond lengths	2	0.935	0.888
	BCP properties	2	0.956	0.927

The values of relevant QCT descriptors ("raw data") of all esters are given as Supporting Information (Table S1). Table 2 shows the summary of seven PLS analyses of ethyl-containing esters. These models contain information from nine bonds and hence entail 9 (level A) or $4 \times 9 = 36$ (levels B, C, and E) descriptors. Excellent statistics are obtained for all ab initio levels of theory (B, C, and E), demonstrating that the extra CPU time compared to that of the semiempirical level A was worth investing. All models passed the randomization validation test and involved a modest two latent variables. The bond length model at level C turned out to be the best with $r^2 = 0.969$ and $q^2 = 0.948$. The three bond lengths that emerged with the highest VIP values were C_1═O_2, C_4—C_5, and O_3—C_4, in that order. One expects to identify the central fragment (O_2═C_1)—O_3—C_4 as the active center because it contains the bond being broken during the hydrolysis and the two neighboring bonds, one at either side of C_1—O_3. Moreover, according to the familiar base-promoted formal reaction mechanism, the C═O bond is most affected by the nucleophilic attack of OH^- to the carbonyl carbon. We recover two of the three expected bonds but are inclined to consider C_4—C_5 as a "contamination". Attempts to eliminate such contaminations have failed so far. One attempt involved the introduction of an extra conformer for each ester and doubling the number of descriptors. Another attempt was to construct a principal component for each bond, consisting

TABLE 3: Survey of the PLS Analyses Obtained for the Complete Set of 40 Esters

level	descriptor	LV	r^2	q^2
A	bond lengths	1	0.715	0.665
B	bond lengths	1	0.737	0.695
	BCP properties	1	0.734	0.682
C	bond lengths	2	0.916	0.899
	BCP properties	3	0.930	0.863
E	bond lengths	2	0.889	0.869
	BCP properties	2	0.899	0.867

of its BCP properties at levels B, C, and E. The subsequent PLS analysis shifted the contamination to another bond. A final comment regarding Table 2 concerns an alternative model that includes only the three bonds in $(O_2{=}C_1){-}O_3{-}C_4$ instead of all nine bonds. With only three bonds model E with BCP properties comes out as the best. This is rewarding because it emphasizes that the more realistic quantum chemical method, which includes electron correlation, improves the model. Furthermore, the added value of the QCT descriptors over and above the mere bond lengths becomes apparent. It is worth pointing out that model E with the BCP properties of all nine bonds is the next best model after that at level C with just bond lengths.

In view of the strong performance of model C it was selected to investigate the subset of 14 acetates. This model, which included 6 bonds (and hence 24 descriptors), yielded an r^2 value of 0.940 and q^2 of 0.865 with two LVs. In the VIP plot the four descriptors of highest VIP value stand out significantly above the others and most importantly involve *only* the C=O bond and the $C_1{-}O_3$ bond. If only bond lengths are allowed as descriptors, the $C_1{-}O_3$ bond, which is broken upon hydrolysis, features again the top VIP value, but the statistics deteriorate somewhat: $r^2 = 0.910$ and $q^2 = 0.781$ (with two LVs). Conclusions remain unaltered if bond and BCP properties are combined in a model of 30 descriptors.

The same analysis was carried out on the subset of 12 benzoate esters, which contain 14 common bonds. The 56-descriptor model (again at level C) yields poorer but still reasonable statistics: $r^2 = 0.842$ and $q^2 = 0.680$ with only one LV. The top four VIPs again refer to $C_1{=}O_2$ and $C_1{-}O_3$. When only bond lengths are included, the statistics change to $r^2 = 0.798$ and $q^2 = 0.772$ (with one LV), favoring $C_1{=}O_2$ as the top VIP. A model combining bond length and BCP properties again highlights the $C_1{=}O_2$ and $C_1{-}O_3$ bonds as most important.

Our analyses of the three largest ester subsets have all pointed to the $(C_1{=}O_2){-}O_3{-}C_4$ fragment as the most important to hydrolysis rate constants. We do not claim that all models *each* include *all* three, but we find that *overall* these bonds invariably score higher than any other bonds in the VIP plots. This observation justifies zooming in on the $(C_1{=}O_2){-}O_3{-}C_4$ fragment and including only its QCT descriptors in the newly constructed models.

Table 3 summarizes the statistics of the models associated with the 12 descriptors of the active center $(C_1{=}O_2){-}O_3{-}C_4$. The model at level C with BCP properties is the best with $r^2 = 0.930$, $q^2 = 0.863$, and three LVs. We observed that the 2,4-D esters greatly benefit from basis set quality because when left out level B performs on par with level C. Figure 2 shows the VIP plot for this model. Clearly, the Laplacian of the electron density at the BCP of $C_1{=}O_2$ is the most significant descriptor to explain the total range of rate constants. This is a meaningful finding in the sense that the Laplacian is a crude measure for bond ionicity, or in a wider context bond polarity. In the classical

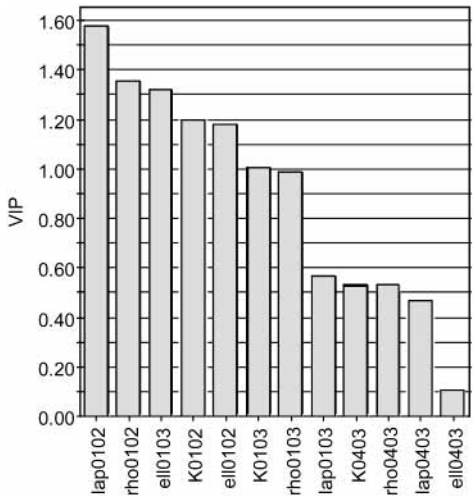

Figure 2. VIP plot for the complete set of 37 esters at level C with BCP properties (QCT descriptors), using the ester group $(C_1{=}O_2){-}O_3{-}C_4$ as the common skeleton. The labels "rho", "lap", "ell", and "K" refer to the electron density, Laplacian, ellipticity, and kinetic energy density at the bond critical point, respectively.

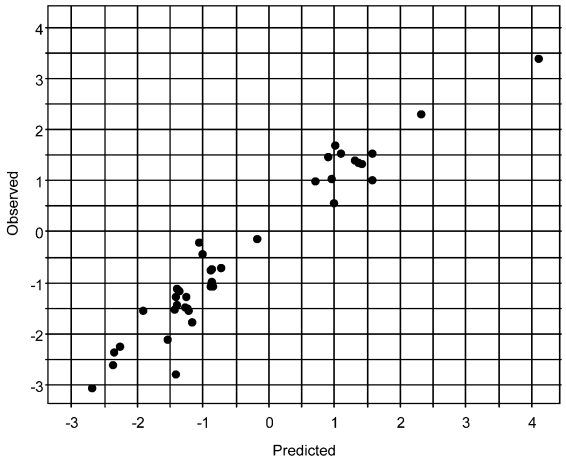

Figure 3. Observed versus predicted rate constant (log k_{OH}) for the complete set of 37 esters at level C with BCP properties (QCT descriptors), using the ester group $(C_1{=}O_2){-}O_3{-}C_4$ as the common skeleton.

base-promoted formal reaction mechanism the OH^- nucleophile attacks the carbonyl group, which is represented by $C^{\delta+}{=}O^{\delta-}$. One would expect the polarity of the $C_1{=}O_2$ bond to affect the propensity for nucleophilic attack, leading to an anionic intermediate with a tetracoordinated carbon. By inspection we confirm the trend that the most positive Laplacian values are found for compounds with the highest rate constant. So, for example, the compound that hydrolyzes fastest, ethyl trichloroacetate, exhibits a Laplacian at the $C_1{=}O_2$ BCP with a value in the upper third. Reassuringly, previous work[17] used gas-phase IR frequencies of exactly the $C_1{=}O_2$ (and $C_1{-}O_3$) bonds to set up an SAR for hydrolysis constants. In that study it was recognized that the polar nature of R strongly affects the receptivity of the nucleophilic attack of OH^- and influences the double-bond character of $C_1{=}O_2$.

Ultimately, Figure 3 shows how well the rate constants are predicted by our final model (i.e., level C, BCP properties of only $(C_1{=}O_2){-}O_3{-}C_4$). We observe that all 40 esters are predicted within 1 log unit. This result improves the previously obtained[17] infrared QSAR, which was confronted with five outliers. That study yielded an r^2 value of 0.887 and failed to report a q^2 or any randomization validation.

In this work we used QTMS slightly differently than in refs 27 and 36 on two counts. First, for the first time we interpret an individual QCT descriptor (i.e., the Laplacian) rather than a principal component, constructed from QCT descriptors, describing a bond. Second, the flexibility of QTMS is further underlined by its "hierarchical" use of common skeletons to highlight an active center.

4. Conclusions

Thanks to the recent abundance of inexpensive computing power, we are able to construct QSARs based on ab initio wave functions. We show that optimized bond lengths and QCT descriptors perform well as novel descriptors in predicting the rate hydrolysis of a set of 40 esters. Models for three subsets, each having a different common skeleton, point toward the three central ester bonds as the active center. A new model that just includes those three bonds links the Laplacian of the electron density at the C═O bond critical point to the formal reaction mechanism of base-promoted ester hydrolysis. This successful application of the QTMS method demonstrates that there is no need to measure IR frequencies, which itself has been introduced as a faster method to avoid the measurement of rate constants. With an eye on current unpublished results, we are confident that the applicability of QTMS extends to the successful prediction of hydrolysis rates of other compounds and reaction parameters such as pK_a.

Acknowledgment. Gratitude is expressed to EPSRC and ICI for sponsoring this work.

References and Notes

(1) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, 1st ed.; American Chemical Society: Washington, DC, 1995; Vol. 1.

(2) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995.

(3) Leo, A. J.; Hansch, C. *Perspect. Drug Discovery Des.* **1999**, *17*, 1.

(4) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027.

(5) Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley-Interscience: New York, 2000.

(6) Fukui, K. *Theory of Orientation and Stereoselection*; Wiley: New York, 1975.

(7) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.

(8) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, Germany, 2000.

(9) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford University Press: Oxford, 1990.

(10) Popelier, P. L. A. *Atoms in Molecules. An Introduction*; Pearson Education: London, 2000.

(11) Popelier, P. L. A.; Smith, P. J. Quantum Topological Atoms. In *Chemical Modelling: Applications and Theory*; Hinchliffe, A., Ed.; Royal Society of Chemistry Specialist Periodical Report; Royal Society of Chemistry: London, 2002; Vol. 2, Chapter 8, pp 391−448.

(12) Popelier, P. L. A.; Aicken, F. M.; O'Brien, S. E. Atoms in Molecules. In *Chemical Modelling: Applications and Theory*; Hinchliffe, A., Ed.; Royal Society of Chemistry Specialist Periodical Report; Royal Society of Chemistry: London, 2002; Vol. 1, Chapter 3, pp 143−198.

(13) Collette, T. W. *TrAC, Trends Anal. Chem.* **1997**, *16*, 224.

(14) Peijnenburg, W. J. G. M. *Pure Appl. Chem.* **1991**, *63*, 1667.

(15) Tratnyek, P. G. In *Perspectives in Environmental Chemistry*; Macalady, D. L., Ed.; Oxford University Press: New York, 1998; p 167.

(16) Wade, L. G. *Organic Chemistry*; Prentice Hall: Upper Saddle River, NJ, 1999.

(17) Collette, T. W. *Environ. Sci. Technol.* **1990**, *24*, 1671.

(18) Popelier, P. L. A.; Chaudry, U.; Smith, P. J. *J. Chem. Soc., Perkin Trans. 2* **2002**, 1231.

(19) O'Brien, S. E.; Popelier, P. L. A. *Quantum Molecular Similarity: Use of Atoms in Molecules derived quantities as QSAR variables*; ECCOMAS: Barcelona, Spain, 2000.

(20) Smith, P. J.; Popelier, P. L. A. Submitted for publication.

(21) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.

(22) Wagener, M.; Sadowski, J.; Gasteiger, J. *J. Am. Chem. Soc.* **1995**, *117*, 7769.

(23) Alsberg, B. K.; Marchand-Geneste, N.; King, R. D. *Chemom. Intell. Lab. Syst.* **2000**, *54*, 75.

(24) Alsberg, B. K.; Marchand-Geneste, N.; King, R. D. *Anal. Chim. Acta* **2001**, *446*, 3.

(25) Platts, J. A. *Phys. Chem. Chem. Phys.* **2000**, *2*, 973.

(26) Platts, J. A. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3115.

(27) O'Brien, S. E.; Popelier, P. L. A. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764.

(28) Silverman, B. D.; Platt, D. E. *J. Med. Chem.* **1996**, *39*, 2129.

(29) Coats, E. A. *Perspectives Drug Discovery Des.* **1998**, *12*, 199.

(30) Carbó-Dorca, R.; Robert, D.; Amat, L.; Girones, X.; Besalu, E. *Molecular Similarity in QSAR and Drug Design*; Springer: Berlin, 2000; Vol. 73.

(31) Carbó, R.; Besalu, E.; Amat, L.; Fradera, X. *J. Math. Chem.* **1995**, *18*, 237.

(32) Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*, 2nd ed.; Gaussian Inc.: Pittsburgh, PA, 1996.

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *GAUSSIAN 98*, Revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(34) O'Brien, S. E.; Popelier, P. L. A. *Can. J. Chem.* **1999**, *77*, 28.

(35) O'Brien, S. E. Quantum Molecular Similarity, An Atoms in Molecules Approach. Ph.D. Thesis, UMIST, Manchester, England, 2000.

(36) O'Brien, S. E.; Popelier, P. L. A. *J. Chem. Soc., Perkin Trans. 2* **2002**, 478.

(37) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(38) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(39) *MORPHY98. A program written by P. L. A. Popelier with a contribution from R. G. A. Bone*; UMIST: Manchester, England, 1998; http:// morphy.ch.umist.ac.uk/.

(40) Wold, S.; Sjostrom, M.; Eriksson, L. Partial Least Squares Projections to Latent Structures (PLS) in Chemistry. In *Encyclopedia of Computational Chemistry*; Schleyer, P.v. R., Ed.; Wiley: New York, 1998; p 2006.

(41) *SIMCA-P 8.0*; UMETRICS: Umea, Sweden, 1998; info@ umetrics.com; www.umetrics.com.

(42) Livingstone, L. *Data Analysis for Chemists*; Oxford University Press: Oxford, 1995.

(43) *SIMCA-P 8.0 User Guide and Tutorial*; UMETRICS: Umea, Sweden, 1999.