ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/264369657

Predictive QSPR modeling of acidic dissociation constant (pKa) of phenols in different solvents

ARTICLE *in* JOURNAL OF PHYSICAL ORGANIC CHEMISTRY · MARCH 2009

Impact Factor: 1.38 · DOI: 10.1002/poc.1447

CITATIONS
26

READS
136

2 AUTHORS:

Kunal Roy
Jadavpur University
293 PUBLICATIONS **5,066** CITATIONS
SEE PROFILE

Paul L A Popelier
The University of Manchester
190 PUBLICATIONS **7,111** CITATIONS
SEE PROFILE

Available from: Paul L A Popelier
Retrieved on: 24 January 2016

Research Article

Journal of Physical Organic Chemistry

Received: 20 May 2008, Revised: 31 July 2008, Accepted: 15 August 2008, Published online in Wiley InterScience: 21 October 2008

(www.interscience.wiley.com) DOI 10.1002/poc.1447

Predictive QSPR modeling of the acidic dissociation constant (pK_a) of phenols in different solvents

Kunal Roy[a]* and Paul L. A. Popelier[b]**

Given the importance of ionization constant (pK_a) of phenols in explaining the mechanism of their toxicity, it is of interest to develop theoretical models for the prediction of pK_a values of phenols in different solvent systems. In the present communication, we developed predictive QSPR models for pK_a values of substituted phenols in seven different solvent systems such as water, dimethyl sulfoxide (DMSO), methanol, dimethylformamide (DMF), acetonitrile (AN), isopropanol, and *tert*-butanol using quantum topological molecular similarity (QTMS) descriptors. The data set was divided into training and test sets, and models were developed using partial least squares (PLS) regression from the training set. The predictive potential of the developed models was assessed by the prediction of pK_a values of the test set compounds. Root mean square error of prediction (RMSEP) values were used as objective function for selection of the best models in different solvent systems. Good predictive models were developed in all solvent systems except *iso*propanol. Considering all seven solvent systems, distance descriptors give consistently good results whereas ellipticity descriptors are of less importance. Moreover, plots of 'variable importance in the projection' (VIP) for the best models highlight the importance of the bond connecting the phenolic oxygen to the aromatic ring. This suggests the diagnostic nature of QTMS descriptors in identifying the reaction center in acidic dissociation of phenols.
Copyright © 2008 John Wiley & Sons, Ltd.

Keywords: QTMS; pK_a; *ab initio*; phenols; QSPR; external validation; electron density; atoms in molecules; quantum chemical topology

INTRODUCTION

Phenolic compounds have commercial uses as intermediates of dyes and organic synthesis processes and also as antiseptics and disinfectants.[1] Phenols are frequently detected in industrial waste water, including those from the manufacture of insecticides, herbicides, dyes, pulp, paper, and other synthetic chemicals.[2] Due to their many origins and widespread uses, phenols are widely distributed in the ecosystem and they have potential to cause environmental pollution.[1] The toxic potency of phenols is controlled by both hydrophobicity and ionization.[3] Various types of phenols also occur in different natural products such as fruits, vegetables, and teas. The phenolic moiety is found in natural radical scavengers (Vitamin E) as well as synthetic ones such as butylated hydroxyanisole.[4] These compounds show a wide action spectrum involving antitumor, antiviral, antibacterial, cardioprotective, prooxidant, and antimutagenic activity.[5] Therefore, a confusing problem with phenols is that there are 'good' phenols and 'bad' phenols. Phenolic compounds may both scavenge and generate reactive oxygen species, and the development of phenolic antioxidants for clinical use includes strategies to minimize the prooxidant activity.[6] This necessitates application of a rational strategy to understand the relationship between structure and functions of phenols.

Quantitative structure–property relationships (QSPRs) represent predictive models derived from the application of statistical tools correlating property (physicochemical property or therapeutic and toxic activity) of chemicals (industrial chemicals/drugs/toxicants/environmental pollutants) with descriptors representative of molecular structure and/or property. A QSPR model is regarded as a scientifically credible method for determining properties of untested chemicals. Any QSPR modeling should ultimately lead to statistically robust models capable of making accurate and reliable predictions of properties of new compounds. QSPRs of toxicity of phenols have been studied by different groups of authors.[1–12] The ionization

* *Correspondence to: K. Roy, Drug Theoretics and Cheminformatics Lab, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India.*
E-mail: kroy@pharma.jdvu.ac.in

** *Correspondence to: P. L. A. Popelier, Manchester Interdisciplinary Biocenter (MIB), 131 Princess Street, Manchester M1 7DN, UK.*
E-mail: pla@manchester.ac.uk

a *K. Roy*
Drug Theoretics and Cheminformatics Lab, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

b *P. Popelier*
Manchester Interdisciplinary Biocenter (MIB), 131 Princess Street, Manchester M1 7DN, UK and School of Chemistry, Univ. of Manchestor, Oxford Road, Manchester, M13 9PL, UK.

Copyright © 2008 John Wiley & Sons, Ltd.

constant (pK_a) of phenols was found to be one of the important contributors of toxicity of phenols.[2,3,5,12] Phenols elicit toxic responses by one of two mechanisms: polar narcosis and uncoupling of oxidative phosphorylation. Polar narcosis has been reported to be the main mechanism of phenol toxicity and it can be modeled well using hydrophobicity and pK_a values of phenols as descriptors.[1] The ability to act as oxidative uncouplers is also associated with pK_a values.[11] Thus, pK_a values are useful in predicting the mechanism of toxic actions of phenols.[12]

Experimentally derived pK_a values are not always available from the literature. Considering the importance of pK_a values in predicting the mechanism of toxic action of phenols, it is of interest to develop theoretical models for prediction of pK_a values of phenols in different solvent systems. Although most experimental values of pK_a of phenols have been determined in water, the importance of nonaqueous solutions is increasing.[13]

QTMS descriptors are known to model well properties and activities for which the electronic factor is important.[14–28] In the present communication, we developed predictive QSPR models for pK_a values of substituted phenols in seven different solvent systems, such as water, dimethyl sulfoxide (DMSO), methanol, dimethylformamide (DMF), acetonitrile (AN), isopropanol, and *t*-butanol.

MATERIALS AND METHODS

The experimental pK_a values of phenols were taken from Reference.[13] Excluding iodine containing compounds (because of the lack of prestored basis sets), a total of 90 phenolic compounds were considered. The experimental pK_a values were available for 62 phenols for water, 35 phenols for DMSO, 31 phenols for methanol, 22 phenols for DMF, 21 phenols for AN, and 13 phenols for each of isopropanol and *t*-butanol (Table 1).

The details of QTMS descriptors can be found in previous publications.[16,21] In summary, QTMS descriptors focus on bond critical points (BCP), which occur when the gradient of the electron density ($\nabla\rho = \mathbf{0}$) vanishes at some point between two bonded nuclei. The electron density at a BCP can be related to bond order via an exponential relationship. At a BCP, the Hessian of ρ has two negative eigenvalues ($\lambda_1 < \lambda_2 < 0$) and one positive one ($\lambda_3 > 0$). Eigenvalues express local curvature of ρ in a point: negative eigenvalues are curvatures perpendicular to the bond, while the positive eigenvalue measures the curvature along the bond. The sum of the eigenvalues is the Laplacian, denoted by $\nabla^2\rho$, which is a measure of how much ρ is concentrated ($\nabla^2\rho < 0$) or depleted ($\nabla^2\rho > 0$) in a point. The descriptors ρ_b and λ_3 can be interpreted as measures of σ character whilst $\lambda_1 + \lambda_2$ measures the degree of π character.[29] Another measure of π character for homopolar bonds is ellipticity which is defined as $\varepsilon = \lambda_1/\lambda_2 - 1$. In the QTMS bond descriptor vector, there are two more components, the kinetic energy density $K(\mathbf{r})$ and a more classical kinetic energy $G(\mathbf{r})$, as defined earlier.[30] Additionally, the equilibrium bond length (R_e) has also been used as one of the descriptors along with other QTMS descriptors.

To start, an estimated geometry was obtained using the program GaussView,[31] which was then passed on to the *ab initio* program GAUSSIAN03.[32] The wavefunctions were computed at B3LYP/6-31 + G(d,p) level of theory. Subsequently, the wavefunctions were read by a local version of the program MORPHY98,[33] which locates the BCPs using an automatic and robust algorithm.[34] The BCP descriptors of eight bonds common to all phenolic compounds (six C—C aromatic bonds, one C—O bond, and one O—H bond) were considered as variables for the statistical model development. The bonds are shown in Fig. 1. Finally, the program SIMCA[35] was used for partial least squares (PLS) analysis of the data set. This procedure generalizes and combines features from principal component and multiple regression.[36,37] The principal component that PLS generates are called latent variables (LV). To avoid overfitting, a strict test for the significance of each consecutive PLS component is necessary and no new LVs are added when they become nonsignificant. This ensures that the QSAR equations are selected based on their ability to predict the data rather than to fit the data. Cross-validation (CV) is a practical and reliable method of testing this significance.[37] With CV, parts of the data are kept out of model development, and then predicted by the model, and compared with the actual values. This procedure is repeated several times until every observation has been kept out once and only once. For every component, the overall PRESS/SS is computed, where PRESS is the squared differences between observed Y and predicted values when the observations were kept out and SS is the residual sum of squares of the previous component. A component is considered significant if PRESS/SS is statistically smaller than 1.0.[35]

For the development of the PLS models, a hierarchical method was adopted. Initially, PLS models were developed for each category of descriptors, i.e., ρ, $\nabla^2\rho$, ε, λ, K, G, and equilibrium bond lengths. Note that λ_1, λ_2, and λ_3 are clubbed together into the class of λ. There are $3 \times 8 = 24$ descriptors in the class of λ while in other classes there are only eight. At the outset, models were tried with all available descriptors, but subsequently, descriptors with smaller VIP (variable importance in the projection) values were gradually deleted until a model with the best Q^2 (leave-one-seventh-out cross-validation) was obtained. Then, using the important descriptors appearing in the PLS equations of different descriptor classes, a PLS model for a mixed set of descriptors was developed. In the case of leave-some-out (LSO) cross-validation, a given fraction of compounds were deleted from the data set and a model was developed from the reduced set. The deleted compounds were predicted from the model, which was developed excluding these compounds. The process was repeated until all compounds were deleted at least once. The outcome from the cross-validation procedure is cross-validated R^2 (LSO-Q^2) which is used as a criterion of both robustness and predictive ability of the model.

External validation is an important tool for proper selection of QSPR models. The validation strategies check the reliability of the developed models for their possible application on a new set of data, and confidence of prediction can thus be judged.[38,39] In many cases, enough new chemicals being unavailable for prediction purpose, the original data set is divided into a training set and a test set. For the present work, every fourth compound of the data set was assigned to the test set. For external validation, a predictive coefficient R^2_{pred} was calculated via Eqn (1)

$$R^2_{pred} = 1 - \frac{\sum (Y_{obs} - Y_{pred})^2}{\sum (Y_{obs} - \overline{Y}_{Training})^2} \quad (1)$$

where Y_{obs} and Y_{pred} respectively represent the observed and predicted property values of the test set compounds, while $\overline{Y}_{Training}$ represents the mean observed value of the training set. The R^2_{pred} value is in part controlled by the magnitude of $\sum (Y_{obs} - \overline{Y}_{training})^2$. This difference is in turn dependent on the

Copyright © 2008 John Wiley & Sons, Ltd.

Table 1. Observed and calculated values of phenols in different solvent systems

		pK_a values of phenols in different solvent systems													
		Water		DMSO		Methanol		DMF		AN		Isopropanol		*t*-butanol	
Sl. no.	Names of phenols	Obs.[a]	Calc.[b]	Obs.[a]	Calc.[c]	Obs.[a]	Calc.[d]	Obs.[a]	Calc.[e]	Obs.[a]	Calc.[f]	Obs.[a]	Calc.[g]	Obs.[a]	Calc.[h]
						Training set									
1	2,3,4,6-Tetrachlorophenol	5.63	7.24	7.55	9.87			9.50	7.99						
2	2,3,4,6-Tetranitrophenol							1.11	2.07						
3	2,3-Dihydroxyphenol	9.01	8.49												
5	2,3-Dinitrophenol	5.24	6.46			9.43	10.03								
6	2,4,5-Trichlorophenol			10.97	12.27			12.46	12.70						
7	2,4,6-Tribromophenol	6.10	6.30			10.10	9.88								
9	2,4,6-Trimethyl-3-nitrophenol	8.98	9.45												
10	2,4,6-Trimethylphenol	10.86	9.98			15.53	14.23								
11	2,4,6-Trinitrophenol	0.43	2.01	−0.30	0.75	3.90	4.42	3.65	2.23	11.00	12.07	3.70	4.70	4.70	5.50
13	2,4-Dimethylphenol	10.6	9.92			15.04	14.96								
14	2,4-Dinitrophenol	4.10	4.37	5.32	5.02	7.82	7.79	6.36	7.86	18.40	17.44				
15	2,4-Di-*tert*-butylphenol	11.57	10.17			16.77	16.34								
17	2,5-Dimethylphenol	10.41	10.34			14.91	15.41								
18	2,5-Dinitrophenol	5.22	4.55	7.32	5.07	8.93	8.87	8.78	8.43						
19	2,6-Dibromo-4-nitrophenol	3.38	4.50	5.17	6.93	7.31	7.69	5.71	6.96						
21	2,6-Dichloro-4-nitrophenol	3.55	4.51			7.40	7.41	5.72	6.96						
22	2,6-Dichlorophenol	6.79	5.92	11.54	9.43			12.55	11.00			13.58	12.73	16.38	15.73
23	2,6-Dimethyl-4-cyanophenol	8.27	7.81												
25	2,6-Dimethylphenol	10.59	9.89			15.26	14.24								
26	2,6-Dinitro-4-hydroxyphenol	4.42	4.26												
27	2,6-Dintrophenol	3.74	3.12	4.82	3.43	7.70	6.73	6.07	4.71	16.45	14.36				
29	2,6-Di-*tert*-butylphenol			16.85	13.30										
30	2-Aminophenol	9.44	10.67												
31	2-Bromophenol	8.39	7.53					13.85	14.02	23.92	23.43				
33	2-Chloro-4-phenylphenol	8.07	7.14	14.90	12.75			15.70	14.43	24.90	23.14	13.70	14.94	18.60	19.92
34	2-Chlorophenol	8.51	7.64			12.83	12.77					15.83	14.67	18.54	18.03
35	2-Ethylphenol	10.20	9.80												
37	2-Hydroxy-3-nitrophenol	6.68	11.26												
38	2-Hydroxyphenol	9.12	8.69												
39	2-Methoxyphenol	9.90	9.15			14.48	14.67								
41	2-Methylphenol	10.31	9.82			14.90	14.92			27.50	26.75				
42	2-Nitrophenol	7.23	6.13	11.00	7.58	11.52	10.70	12.14	11.93	22.00	22.25	13.30	12.29	15.88	14.11
43	2-*tert*-butylphenol	11.34	9.84			16.50	15.75								
45	3-(Trifluoromethylsulfonyl)phenol														
46	3,4,5-Trichlorophenol	7.68	7.94	12.58	14.32										
47	3,4-Dichlorophenol	8.51	8.75	14.22	14.30			13.22	13.16	24.06	23.58				

www.interscience.wiley.com/journal/poc

Copyright © 2008 John Wiley & Sons, Ltd.

49	3,4-Dinitrophenol	5.42	6.54	7.97	11.40	9.46	9.76			17.90	18.71				
50	3,5-Dichlorophenol	8.18	7.73	13.09	12.58	12.94	11.88			23.31	21.74	14.05	14.84	17.04	17.03
51	3,5-Dihydroxyphenol	8.45	8.05												
53	3,5-Dimethyl-4-nitrophenol	8.25	9.15												
54	3,5-Dimethylphenol	10.20	10.36			14.62	14.94								
55	3,5-Dinitrophenol	6.66	7.19	10.60	11.52	10.20	11.59	11.30	10.99	20.50	21.74	10.84	10.29	13.40	12.60
57	3-Acetylphenol	9.19	10.15	15.14	15.85										
58	3-Aminophenol	9.99	10.19												
59	3-Bromophenol	9.01	8.84			13.30	13.05					14.83	16.13	18.52	19.37
61	3-Chloro-4-nitrophenol	6.49	6.94	9.80	12.22					19.95	20.79				
62	3-Chlorophenol	9.02	8.84	15.83	14.33	13.10	13.17			25.04	24.93				
63	3-Cyanophenol			14.76	14.94										
65	3-Fluorophenol	9.28	7.87	15.88	13.56										
66	3-Hydroxyphenol	9.15	8.64	15.30	14.56										
67	3-Methoxyphenol	9.65	8.83	15.72	14.92										
69	3-Methylphenol	10.10	9.89	16.86	15.97	14.48	15.86								
70	3-Nitrophenol	8.36	8.50	13.75	13.89	12.40	12.73	13.85	15.04	23.85	24.67	13.92	14.09	16.99	17.46
71	3-Trifluoromethyl-4-nitrophenol	6.41	7.79	9.30	12.81			10.40	11.02	19.30	21.26	9.90	10.48	12.77	14.18
73	4-Acetylphenol	8.05	7.79	13.68	13.91										
74	4-Aminophenol	10.43	10.04												
75	4-Bromophenol	9.36	9.56	15.50	14.93	13.63	13.69	14.34	14.42	25.53	24.86	14.30	14.67	19.1	18.54
77	4-Chlorophenol	9.38	9.62	16.10	14.88	13.59	13.86	14.50	14.02	25.44	25.18	15.31	14.50	18.96	18.32
78	4-Cyanophenol	7.8	7.36	13.01	13.27					22.77	23.65				
79	4-Ethylphenol	10.00	9.68												
81	4-Hydroxy-2,3,5,6-tetramethylphenol	11.51	12.31												
82	4-Hydroxy-2-methylphenol	10.20	10.69												
83	4-Hydroxy-2-nitrophenol	7.63	6.78												
85	4-Methoxyphenol	10.27	10.53	17.58	16.63										
86	4-Methylphenol	10.28	9.64	16.96	15.85	14.52	14.76			27.45	27.40				
87	4-Nitrophenol	7.18	7.45	11.00	12.14	11.24	11.27	11.84	11.49	20.70	21.58	12.45	11.37	14.60	14.68
89	Pentachlorophenol			7.05	12.82			7.97	9.72						
90	Phenol	9.99	9.57	16.47	15.28	14.32	14.76	15.40	15.27	26.60	27.05				
	Average (training set)	8.23		12.09		12.20		10.29		22.22		12.75		15.81	
	Median (training set)	8.75		13.09		13.10		11.57		23.31		13.70		16.99	
	Maximum (training set)	11.57		17.58		16.77		15.70		27.50		15.83		19.10	
	Minimum (training set)	0.43		−0.30		3.90		1.11		11.00		3.70		4.70	
						Test set									
4	2,3-Dimethylphenol	10.54	10.57			15.08	15.74								
8	2,4,6-Trichlorophenol	6.42	6.45	10.19	9.71			12.05	9.66			12.55	12.42	14.82	15.37
12	2,4-Dichlorophenol	7.65	7.95	13.25	12.32			13.56	12.85			14.48	14.16	17.25	17.32
16	2,5-Dichlorophenol							13.85	12.42						
20	2,6-Dichloro-4-hydroxyphenol	7.38	6.80												
24	2,6-Dimethyl-4-nitrophenol	7.19	7.85	9.00	11.43										
28	2,6-Di-*tert*-butyl-4-nitrophenol	6.62	7.11	7.60	7.99	10.89	9.55	8.27	8.38	19.10	18.57				
32	2-Chloro-4-bromophenol	7.64	7.86			12.70	12.13								

(*Continues*)

Copyright © 2008 John Wiley & Sons, Ltd.

Table 1. (Continued)

Sl. no.	Names of phenols	Water		DMSO		Methanol		DMF		AN		Isopropanol		*t*-butanol	
		pK_a values of phenols in different solvent systems													
		Obs.[a]	Calc.[b]	Obs.[a]	Calc.[c]	Obs.[a]	Calc.[d]	Obs.[a]	Calc.[e]	Obs.[a]	Calc.[f]	Obs.[a]	Calc.[g]	Obs.[a]	Calc.[h]
		Training set													
36	2-Fluorophenol	8.73	7.12			12.14	12.19								
40	2-Methyl-4,6-dintrophenol			4.59	4.61										
44	3-(Methylsulfonyl)phenol			13.56	14.48										
48	3,4-Dimethylphenol	10.36	9.96			14.63	16.37								
52	3,5-Dimethyl-4-cyanophenol	8.21	8.21												
56	3,5-Di-*tert*-butylphenol	10.29	10.83			14.89	15.48								
60	3-Chloro-2,4,6-trinitrophenol			1.16	2.14										
64	3-Ethylphenol	9.90	9.88												
68	3-Methyl-5-ethylphenol	10.10	10.48												
72	3-Trifluoromethylphenol	9.04	9.33	14.30	14.73	12.10	13.56	15.70	14.59	24.90	25.56	12.50	13.67	17.10	17.59
76	4-Chloro-2,6-dinitrophenol	2.97	3.39	3.51	3.75			4.68	2.65	15.30	13.69				
80	4-Fluorophenol	9.95	10.33												
84	4-Hydroxyphenol	9.14	10.17												
88	4-*tert*-Butylphenol	10.31	9.72			14.52	14.72			27.48	27.46				
	Average (test set)	8.47		8.57		13.37		11.35		21.70		13.18		16.39	
	Median (test set)	8.89		9.00		13.61		12.81		22.00		12.55		17.10	
	Maximum (test set)	10.54		14.30		15.08		15.70		27.48		14.48		17.25	
	Minimum (test set)	2.97		1.16		10.89		4.68		15.30		12.50		14.82	
	Average (total set)	8.28		11.37		12.44		10.52		22.13		12.83		15.92	
	Median (total set)	8.86		12.80		13.10		11.95		23.31		13.64		17.02	
	Maximum (total set)	11.57		17.58		16.77		15.70		27.50		15.83		19.10	
	Minimum (total set)	0.43		−0.30		3.90		1.11		11.00		3.70		4.70	

[a] Taken from Reference [13].
[b] From model W6.
[c] From model D6.
[d] From model M7.
[e] From model F5.
[f] From model A7.
[g] From model I5.
[h] From model B8.

Copyright © 2008 John Wiley & Sons, Ltd.

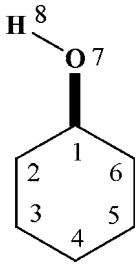

Figure 1. Numbering of atoms in the phenol skeleton

selection of training set members. Thus, R^2_{pred} may not truly reflect the models' predictive capability for the test set (or a new data set) since R^2_{pred} depends on the training set members. So, squared correlation coefficient values between the observed and predicted values of the test set compounds with intercept (r^2) and without intercept (r^2_0) can be calculated to assess performance of the prediction. The parameter r'^2_0 is the squared correlation coefficient between the predicted (Y) and observed (X) values of the test set compounds setting intercept to 0. According to Golbraikh and Tropsha[40] models are considered acceptable, if they satisfy all of the following conditions: (i) $Q^2 > 0.5$, (ii) $r^2 > 0.6$, (iii) r^2_0 or r'^2_0 is close to r^2, such that $[(r^2 - r^2_0)/r^2]$ or $[(r^2 - r'^2_0)/r^2] < 0.1$ and $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$. When the observed values of the test set compounds (Y-axis) are plotted against the predicted values of the compounds (X-axis) setting intercept to 0, slope of the fitted line gives the value of k. Interchange of the axes gives the value of k′.

Moreover, the squared regression coefficient (r^2) between observed and predicted values of the test set compounds does not necessarily indicate that the predicted values are very near to observed property values (there may be considerable numerical differences between the values in spite of maintaining an good overall intercorrelation). To better gauge the external predictive capacity of a model, a modified r^2 term (r^2_m), defined before,[41] is given in Eqn (2)

$$r^2_m = r^2(1 - \sqrt{r^2 - r^2_0}) \quad (2)$$

Note that r^2 is always larger than r^2_0. In case of good external prediction, predicted values will be very close to observed property values. So, the r^2 value will be very near to the r^2_o value, and in the best case, r^2_m will be equal to r^2. For some solvent systems, the number of test set compounds is considerably smaller thus making regression based external validation parameters (R^2_{pred}, r^2, r^2_m) less appropriate. Thus, an additional parameter root mean square error of prediction (RMSEP) was calculated according to Eqn (3)

$$\text{RMSEP} = \sqrt{\frac{\sum(Y_{obs} - Y_{pred})^2}{N_{Test}}} \quad (3)$$

In the above equation, N_{Test} indicates the number of test set compounds.

For all the developed models, we have reported the coefficient of variation (R^2), leave-one-seventh-out cross-validation R^2 (Q^2) for the training set and the R^2_{pred}, r^2, r^2_0, r^2_m and RMSEP values for the test set. The final models were also subjected to a randomization test. In this test, the property data (Y) are randomly permuted keeping the descriptor matrix intact, followed by a PLS run. Each randomization and subsequent PLS analysis generates a new set of R^2 and Q^2 values, which are plotted against the correlation coefficient between the original Y values and the permuted Y values. The intercepts for the R^2 and Q^2 lines in this plot are a measure of the overfit. A model is considered[42] valid if $R^2_{int} < 0.4$ and $Q^2_{int} < 0.05$.

RESULTS AND DISCUSSION

RMSEP values were used as objective function for selection of the best models in different solvent systems. The statistical qualities of different models are shown in Table 2.

QSPR of pK_a values of phenols in water

Acidic dissociation constant (pK_a) values in water were available for 80 phenols among which 62 phenols were assigned to the training set and 18 phenols to the test set. Model W1, developed from eight distance descriptors and two LVs, showed 76% predicted variance (leave-one-seventh out cross-validation) while the R^2 value was 0.825. In order to properly judge the reliability of a model we rate external validation parameters higher than the internal prediction statistics. The application of model W1 to the test set led to a predictive R^2 value of 88.4%. The squared correlation coefficient (r^2) between the observed and predicted values of the test set compounds was found to be 0.912 while the same setting intercept to zero was 0.883, reflecting a moderate decrease in the value of r^2_m to 0.757. The RMSEP value of model W1 was 0.651. The RMSEP values of models W2-W5 (based on ρ, ∇^2_ρ, ε, and λ descriptors) were greater than that of model W1. Model W6, which is based on the K descriptors, showed RMSEP value of 0.584 and R^2_{pred} and r^2_m values of 0.906 and 0.880, respectively. Model W7, based on G descriptors, was slightly inferior to model W6. Based on descriptors appearing in models W1 through W7, model W8 was developed, which was slightly inferior to both models W6 and W7. Thus, based on external validation statistics, model W6 was the best one for predicting reliable pK_a values of phenols in water. The ellipticity based model (W4) showed poor internal and external validation statistics. The VIP plot (Fig. 2) of model W6 showed K descriptor of the bond C1—O7 (bond connecting phenolic oxygen to aromatic nucleus) as the most important descriptor, which is justified considering the dissociation constant of the phenolic compounds as the response parameter. In earlier work,[43] a QTMS study on a set of 19 singly substituted *para* and *meta*-phenols highlighted the O—H bond as having the highest VIP values when looking at the principal components of the descriptors, although the C—O bond featured prominently as well. When the 'raw' descriptors were considered then the four highest VIP values were found to be a mixture of C—O and O—H descriptors. Although O—H is expected to be the most affected by the acid dissociation, C—O is the next most sensitive. Their relationship appears to be too tight to be clearly separable by the statistical analysis we use.

QSPR of pK_a values of phenols in DMSO

Acidic dissociation constant (pK_a) values in DMSO were available for 44 phenols among which 35 phenols were assigned to the training set and nine phenols to the test set. Models were developed from the training set and test set prediction was done to check validity of the developed models. The best model based on internal validation was model D5 developed from 14 λ

Copyright © 2008 John Wiley & Sons, Ltd.

Table 2. Comparative analysis of QSPR models for pK_a of phenols in different solvent systems developed with QTMS descriptors calculated at B3LYP/6-31 + G(d,p) level

Solvent system	Model no.	N_{Train}/N_{Test}	Type of descriptors	No. of descriptors	LV[a]	R^2	Q^2	R^2_{pred}	r^2	r^2_0	r^2_m	RMSEP
Water	**W8**	62/18	**Mixed**	**2**	**1**	**0.829**	**0.827**	**0.899**	**0.914**	**0.898**	**0.798**	**0.607**
	W7		**G**	**6**	**1**	**0.827**	**0.802**	**0.890**	**0.898**	**0.896**	**0.858**	**0.632**
	W6*		**K**	**6**	**2**	**0.826**	**0.785**	**0.906**	**0.909**	**0.908**	**0.880**	**0.584**
	W5		λ	21	2	0.905	0.854	0.701	0.733	0.696	0.592	1.044
	W4		ε	3	2	0.493	0.416	0.224	0.341	0.237	0.231	1.680
	W3		∇^2_ρ	7	2	0.761	0.667	0.759	0.783	0.781	0.748	0.936
	W2		ρ	8	2	0.840	0.782	0.868	0.904	0.870	0.737	0.694
	W1		**Distance**	**8**	**2**	**0.825**	**0.760**	**0.884**	**0.912**	**0.883**	**0.757**	**0.651**
DMSO	**D8**	35/9	**Mixed**	**11**	**2**	**0.896**	**0.853**	**0.933**	**0.914**	**0.911**	**0.864**	**1.470**
	D7		**G**	**7**	**2**	**0.895**	**0.867**	**0.929**	**0.916**	**0.885**	**0.755**	**1.516**
	D6*		**K**	**6**	**1**	**0.787**	**0.764**	**0.968**	**0.959**	**0.957**	**0.916**	**1.011**
	D5		λ	14	2	0.930	0.887	0.859	0.860	0.834	0.721	2.139
	D4		ε	3	2	0.383	0.216	0.476	0.476	0.434	0.378	4.119
	D3		∇^2_ρ	3	1	0.611	0.505	0.853	0.803	0.799	0.752	2.181
	D2		ρ	5	2	0.838	0.766	0.884	0.885	0.883	0.845	1.934
	D1		**Distance**	**4**	**3**	**0.881**	**0.831**	**0.946**	**0.920**	**0.916**	**0.862**	**1.321**
Methanol	M8	31/8	Mixed	13	2	0.957	0.931	0.266	0.500	0.067	0.171	1.627
	M7*		**G**	**7**	**3**	**0.961**	**0.938**	**0.718**	**0.868**	**0.642**	**0.455**	**1.009**
	M6		**K**	**5**	**2**	**0.952**	**0.930**	**0.524**	**0.796**	**0.468**	**0.340**	**1.311**
	M5		λ	8	2	0.969	0.956	0.377	0.661	0.433	0.346	1.498
	M4		ε	5	2	0.664	0.496	0.192	0.076	−0.301	0.029	1.707
	M3		∇^2_ρ	8	1	0.859	0.833	0.055	0.500	0.233	0.242	1.845
	M2		**ρ**	**7**	**3**	**0.962**	**0.935**	**0.404**	**0.754**	**0.339**	**0.268**	**1.466**
	M1		**Distance**	**7**	**2**	**0.959**	**0.944**	**0.502**	**0.835**	**0.441**	**0.311**	**1.340**
DMF	F8	22/6	Mixed	3	2	0.843	0.831	0.809	0.873	0.800	0.637	1.704
	F7		G	4	3	0.892	0.849	0.745	0.764	0.726	0.615	1.971
	F6		**K**	**5**	**1**	**0.787**	**0.763**	**0.824**	**0.974**	**0.960**	**0.859**	**1.635**
	F5*		**λ**	**13**	**2**	**0.937**	**0.884**	**0.850**	**0.955**	**0.922**	**0.782**	**1.508**
	F4		ε	2	1	0.218	0.181	−0.165	0.009	−0.199	0.005	4.210
	F3		∇^2_ρ	3	1	0.725	0.687	0.602	0.744	0.737	0.682	2.459
	F2		**ρ**	**5**	**1**	**0.767**	**0.743**	**0.817**	**0.976**	**0.971**	**0.907**	**1.670**
	F1		**Distance**	**4**	**3**	**0.900**	**0.841**	**0.826**	**0.926**	**0.926**	**0.926**	**1.629**
AN	A8	21/4	Mixed	16	2	0.885	0.849	0.691	0.839	0.717	0.546	2.672
	A7*		**G**	**6**	**2**	**0.931**	**0.901**	**0.964**	**0.994**	**0.965**	**0.825**	**0.912**
	A6		**K**	**8**	**2**	**0.946**	**0.883**	**0.893**	**0.942**	**0.942**	**0.942**	**1.575**
	A5		λ	14	2	0.961	0.929	0.511	0.659	0.650	0.596	3.361
	A4		ε	4	2	0.676	0.526	0.642	0.641	0.641	0.641	2.878
	A3		∇^2_ρ	8	2	0.916	0.832	0.684	0.935	0.857	0.674	2.703
	A2		**ρ**	**7**	**3**	**0.955**	**0.903**	**0.844**	**0.950**	**0.928**	**0.809**	**1.898**
	A1		**Distance**	**8**	**4**	**0.972**	**0.935**	**0.948**	**0.986**	**0.986**	**0.986**	**1.101**
Isopropanol	**I8**	13/3	**Mixed**	**2**	**1**	**0.825**	**0.682**	**−0.064**	**0.831**	**0.831**	**0.831**	**1.049**
	I7		G	3	1	0.756	0.609	−0.966	0.098	−1.225	−0.015	1.426

(*Continues*)

Copyright © 2008 John Wiley & Sons, Ltd.

Table 2. (Continued)

Solvent system	Model no.	N_{Train}/N_{Test}	Type of descriptors	No. of descriptors	LV[a]	R^2	Q^2	R^2_{pred}	r^2	r^2_0	r^2_m	RMSEP
	I6		K	5	2	0.809	0.643	−1.471	0.006	−1.996	−0.002	1.599
	I5*		λ	**11**	**2**	**0.913**	**0.774**	**0.518**	**0.489**	**0.483**	**0.451**	**0.706**
	I4		ε	5	1	0.431	0.088	−3.774	0.553	−1.057	−0.149	2.222
	I3		∇^2_ρ	**3**	**1**	**0.788**	**0.484**	**0.078**	**0.179**	**0.062**	**0.118**	**0.977**
	I2		ρ	4	2	0.856	0.722	−57.168	0.000	−0.787	0.000	1.258
	I1		**Distance**	**5**	**4**	**0.940**	**0.808**	**−0.079**	**0.179**	**0.062**	**0.118**	**1.057**
t-Butanol	**B8***	13/3	**Mixed**	**17**	**2**	**0.944**	**0.884**	**0.885**	**0.973**	**0.960**	**0.862**	**0.425**
	B7		G	6	2	0.880	0.788	−0.256	0.859	−0.590	−0.175	1.408
	B6		K	5	2	0.891	0.834	0.383	0.783	0.213	0.192	0.986
	B5		λ	**14**	**2**	**0.958**	**0.861**	**0.983**	**0.950**	**0.933**	**0.826**	**0.614**
	B4		ε	5	1	0.478	0.142	−8.110	0.030	−5.508	−0.041	3.790
	B3		∇^2_ρ	**5**	**2**	**0.892**	**0.735**	**0.821**	**0.919**	**0.840**	**0.661**	**0.532**
	B2		ρ	5	2	0.934	0.892	0.753	0.745	0.687	0.566	0.624
	B1		**Distance**	**5**	**4**	**0.988**	**0.973**	**0.800**	**0.962**	**0.957**	**0.894**	**0.562**

Bold models indicate the best four models (according to RMSEP values) for different solvent systems.
Starred models indicate the best models (according to RMSEP values) for different solvent systems.

parameters and two LVs showing a Q^2 value of 0.887. Based on external validation (RMSEP value), model D6 developed from six *K* parameters and one LV was the best one showing excellent R^2_{pred} (0.968) and r^2_m (0.916) values. The next best model based on the RMSEP value was mixed model D8 based on 11 descriptors and two LVs. The R^2_{pred} and r^2_m values of this model were 0.933 and 0.864, respectively. The ellipticity based model (D4) showed poor internal and external validation statistics. The VIP plot (Fig. 2) for model D6 showed the K descriptor of C1—O7 bond again as the most important variable.

QSPR of pK_a values of phenols in methanol

Data were available for 39 compounds out of which 31 were taken to form the training set and eight compounds the test set. The regression based external validation parameters (especially r^2_m) were poorer than those of the water and DMSO systems. However, the RMSEP values were in similar range. The best model according to RMSEP was M7, which is based on seven G descriptors and three LVs showing Q^2 and R^2_{pred} values of 0.938 and 0.718, respectively. However, because of the considerable difference between r^2 and r^2_0 values, the r^2_m value was comparatively smaller (less than 0.5) reflecting the difference in observed and predicted values of the test set compounds. However, considering the small size of the test set, the RMSEP value is a more reliable statistic here for external validation than the regression based parameters. The next best model was M6 with five K descriptors and two LVs showing a R^2_{pred} value of 0.524. However, the r^2_m parameter value for this model was very low. As for the water and DMSO systems, the ellipticity based model (M4) showed poor internal and external validation statistics. The VIP plot (Fig. 2) for model M7 showed that the most important variable is the G variable of the C1—O7 bond. This bond was highlighted before in the VIP plots of models W6 and D6.

QSPR of pK_a values of phenols in DMF

Of the available 28 compounds, 22 compounds were assigned to the training set and six compounds to the test set. Based on RMSEP values, model F5 developed from 13 λ descriptors and two LVs was found to be the best model. However, the RMSEP value of model F5 was comparatively larger than the best models of the water, DMSO and methanol systems (models W6, D6 and M7, respectively). The predictive R^2 value for model F5 was 0.850 while the r^2_m value was 0.782. The next best model was model F1 based on distance descriptors showing a higher value of r^2_m (0.926) than model F5. As in the previous three systems, the ellipticity based model (F4) showed poor internal and external validation statistics. According to the VIP plot (Fig. 2) of model F5, the most significant descriptor was λ_3 of the C1—O7 bond, as before.

QSPR of pK_a values of phenols in AN

Out of 25 available compounds, 21 were assigned to the training set and four compounds to the test set. Based on RMSEP values, the best model was A7 comprising of six G descriptors and two LVs. Model A7 showed excellent values of predictive R^2 and r^2_m statistics. The next best model was A1 based on distance descriptors having excellent Q^2 and r^2_m values (0.935 and 0.986, respectively). Surprisingly, model A4 based on ellipticity descriptors showed an acceptable value of predictive R^2 in this

Copyright © 2008 John Wiley & Sons, Ltd.

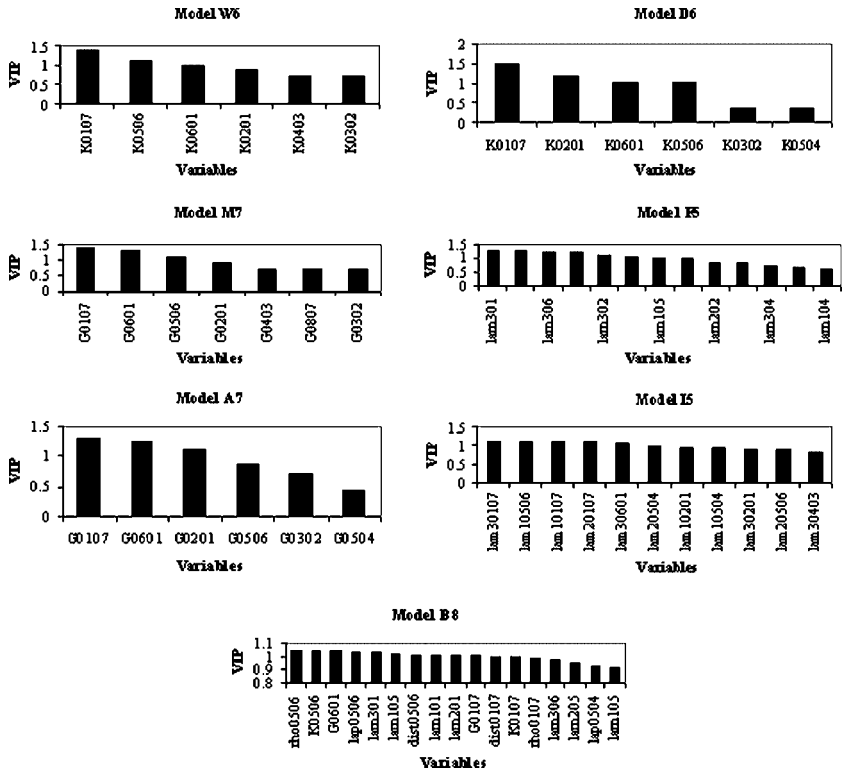

Figure 2. VIP plots for the best models for solvent (a) water W6; (b) DMSO D6; (c) methanol M7; (d) dimethylformamide F5; (e) acetonitrile A7; (f) isopropanol I5; and (g) *t*-butanol B8

solvent system. The VIP plot (Fig. 2) of model A7 showed that the most important variable was G descriptor of the C1—O7 bond.

QSPR of pK_a values of phenols in isopropanol

The data were available only for 16 compounds out of which 13 compounds were taken to the training set and three compounds to the test set. Though the number of available compounds was considerably smaller, a preliminary attempt was made to develop QSAR models out of which model I5 (built by 11 λ descriptors and two LVs) emerged as the best one based on the RMSEP values. However, the r^2_m value of model I5 was below 0.5 and the predictive R^2 value was just above 0.5. The ellipticity based model (I4) was miserably poor in both internal and external validation. The VIP plot (Fig. 2) of model I5 showed λ_3 of the C1—O7 bond as the most important descriptor, as found before for all other solvents.

QSPR of pK_a values of phenols in t-butanol

Among 16 available compounds, 13 compounds were taken to the training set and three compounds to the test set. Based on RMSEP values, model B8 developed from 17 mixed descriptors and two LVs emerged as the best one, showing a R^2_{pred} value of 0.885 and a r^2_m value of 0.862. The next best model was B3 based on Laplacian values showing a predictive R^2 value of 0.821. The r^2_m value of this model was considerably lower (0.661). The ellipticity based model (B4) was wretchedly poor in both internal

Table 3. Occurrence of models of different descriptor types for the best four models of each of different solvent systems

Type of descriptors	Water	DMSO	Methanol	DMF	AN	Isopropanol	*t*-butanol	Total occurrence
Mixed	Y	Y				Y	Y	4
G	Y	Y	Y		Y			4
K	Y	Y	Y	Y	Y			4
λ				Y		Y	Y	3
ε								0
∇^2_ρ						Y	Y	2
ρ			Y	Y	Y			3
Distance	Y	Y	Y	Y	Y	Y	Y	7

'Y' indicates presence of a model of particular descriptor type in the best four models for different solvent systems.

Copyright © 2008 John Wiley & Sons, Ltd.

Table 4. List of R^2_{int} and Q^2_{int} values from randomization test of selected models for pK_a of phenols in different solvent systems

Model no.	R^2_{int}	Q^2_{int}
W6	0.005	−0.221
D6	0.008	−0.181
M7	0.066	−0.440
F5	0.166	−0.300
A7	0.057	−0.366
I5	0.354	−0.254
B8	0.232	−0.286

and external validation. Unlike all other systems, the VIP plot (Fig. 2) of model B8 showed the ρ descriptor of the C5—C6 bond (Fig. 1) as the most important descriptor.

OVERVIEW

The best models in different solvent systems were selected based on the RMSEP values of the corresponding test set compounds. For each solvent we considered the best four models (Table 2). Table 3 shows the number of times a model occurs constructed from each descriptor type. Distance descriptors are among the best four models for all seven solvent systems. However, distance descriptor based models never appear as the best model for a particular solvent. Each of the K, G and mixed descriptor based models appear in four out of seven solvent systems. The ellipticity based models do not appear in the best four models in any solvent system. Thus it may be concluded that considering all seven solvent systems, distance descriptors give consistently good results whereas ellipticity descriptors are of less importance.

Solute–solvent interactions involve charge–dipole, dipole–dipole, dipole–induced dipole, and induced dipole–induced dipole interactions, which are a function of both solute and solvent properties. Different solvent properties like hydrogen bond donation ability, electron pair donation ability, polarizability, dipole moment etc. of solvent are important in determining such interactions. When developing a unified QSAR model for solutes in all solvents, one should consider both solute and solvent properties[44]. Here we have considered the BCP properties of solutes (phenols) only while developing the QSAR models since it is very hard to develop a unified model for all solvent systems without considering the solvent properties. It is also difficult to predict the kind of interactions between a phenol and a particular solvent without considering the solvent parameters. However, the importance of different BCP descriptors for different solvents suggests involvement of different interactions between phenol and solvent depending on the solvent properties. However, future work may include the solvent explicitly by incorporating BCP properties of solute–solvent van der Waals complexes. Jover *et al.* [13] developed a unified neural network model for predicting pK_a of phenols in different solvent systems and they found that the hydrogen bond donation ability and dipole moment of the solvent are important in the development of a unified model for pK_a of phenols in different solvents.

In acidic dissociation of phenols, the phenolic O—H bond is broken and this phenomenon is influenced by the electron density around the oxygen atom. In case of substituted phenols, the change in the electron density around the phenolic oxygen atom is caused by varying substituents on the aromatic ring through polar (resonance and inductive) effects. These effects are transmitted from the aromatic ring to the phenolic oxygen via the C—O bond using a conjugated system (double bond—single bond—lone pair). Thus, the electronic environment of the phenolic oxygen atom is immediately determined by the character of the C—O bond. Interestingly, the VIP plots of the best models of all solvent systems (except *t*-butanol) suggest different descriptors of the phenolic C1—O7 bond as the most important variables. This signifies the diagnostic feature of the QTMS descriptors in identifying important feature (C—O bond as shown in bold line in Fig. 1) for the acidic dissociation of phenols.

The best models of different solvent systems were subjected to a randomization test with 100 permutations (default [35] is 20) in each case. For all the models tested, R^2_{int} values are less than 0.4 and the Q^2_{int} values less than 0.05 (Table 4). This indicates that the models are not obtained by chance. External validation parameters of different models for pK_a values of phenols in different solvent systems according to Golbraikh and Tropsha[38] are shown in Table 5

Table 5. External validation characteristics of different models for pK_a values of phenols in different solvent systems according to Golbraikh and Tropsha[38]

Statistical parameters		Model number						
Sl. no.	Parameters	W6	D6	M7	F5	A7	I5	B8
1	r^2	0.909	0.959	0.868	0.955	0.994	0.489	0.973
2	r_0^2	0.908	0.957	0.642	0.922	0.965	0.483	0.96
3	r'^2_0	0.906	0.953	0.818	0.941	0.974	0.151	0.947
4	$(r^2-r_0^2)/r^2$	0.001	0.002	0.260	0.035	0.029	0.012	0.013
5	$(r^2-r'^2_0)/r^2$	0.003	0.006	0.058	0.015	0.020	0.691	0.027
6	Minimum of 4 and 5	0.001	0.002	0.058	0.015	0.020	0.012	0.013
7	k	0.988	0.959	0.966	1.100	1.008	0.982	0.978
8	k′	1.008	1.034	1.03	0.902	0.99	1.016	1.022

Copyright © 2008 John Wiley & Sons, Ltd.

The calculated pK_a values of phenols in different solvent systems according to the best models are shown in Table 1.

CONCLUSION

Considering all seven solvent systems, distance descriptors give consistently good results whereas ellipticity descriptors are of less importance. The quality of the models was assessed by means of the RMSEP values of an *external* test set. The diagnostic nature of the QTMS descriptors could identify the bond connecting the phenolic oxygen to the aromatic ring as the most important feature for the acidic dissociation of phenols.

Acknowledgements

We thank the Commonwealth Scholarship Commission in the UK for their financial support toward an Academic Staff Fellowship for one of the authors (K.R.).

REFERENCES

[1] X. Wang, J. Yu, Y. Wang, L. Wang, *Chemosphere* **2002**, *46*, 241–250.
[2] Y. G. Lee, S. H. Hwang, S. D. Kim, *Arch. Environ. Contam. Toxicol.* **2006**, *50*, 213–219.
[3] T. W. Schultz, A. P. Bearden, J. S. Jaworska, *SAR QSAR Environ. Res.* **1996**, *5*, 99–112.
[4] R. Garg, A. Kurup, C. Hansch, *Crit. Rev. Toxicol.* **2001**, *31*, 223–245.
[5] M. Y. Moridani, A. Siraki, P. J. O'Brien, *Chem. Biol. Interact.* **2003**, *145*, 213–223.
[6] G. Schüürmann, A. O. Aptula, R. Kuhne, R.-U. Ebert, *Chem. Res. Toxicol.* **2003**, *16*, 974–987.
[7] R. P. Verma, S. Kapur, O. Barberena, A. Shusterman, C. Hansch, C. D. Selassie, *Chem. Res. Toxicol.* **2003**, *16*, 276–284.
[8] C. D. Selassie, T. V. DeSoyza, M. Rosario, H. Gao, C. Hansch, *Chem. Biol. Interact.* **1998**, *113*, 175–190.
[9] G. Schüürmann, H. Segner, K. Jung, *Aquatic Toxicol.* **1997**, *38*, 277–296.
[10] M. T. D. Cronin, A. O. Aptula, J. C. Duffy, T. I. Netzeva, P. H. Rowe, I. V. Valkova, T. W. Schultz, *Chemosphere* **2002**, *49*, 1201–1221.
[11] A. O. Aptula, T. I. Netzeva, I. V. Valkova, M. T. D. Cronin, T. W. Schultz, R. Kuhne, G. Schüürmann, *Quant. Struct. -Act. Relat.* **2002**, *21*, 12–22.
[12] T. W. Schultz, *Ecotoxicol. Environ Saf.* **1987**, *14*, 178–183.
[13] J. Jover, R. Bosque, J. Sales, *QSAR Comb. Sci.* **2007**, *26*, 385–397.
[14] P. L. A. Popelier, *J. Phys. Chem. A* **1999**, *103*, 2883–2890.
[15] S. E. O'Brien, P. L. A. Popelier, *Can. J. Chem.* **1999**, *77*, 28–36.
[16] S. E. O'Brien, P. L. A. Popelier, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764–775.
[17] P. L. A. Popelier, U. A. Chaudry, P. J. Smith, *J. Chem. Soc., Perkin Trans. 2* **2002**, 1231–1237.
[18] S. E. O'Brien, P. L. A. Popelier, *J. Chem. Soc., Perkin Trans.* **2002**, *2*, 478–483.
[19] U. A. Chaudry, P. L. A. Popelier, *J. Phys. Chem. A* **2003**, *107*, 4578–4582.
[20] P. J. Smith, P. L. A. Popelier, *J. Comput. -Aid. Mol. Des.* **2004**, *18*, 135–143.
[21] P. L. A. Popelier, P. J. Smith, U. A. Chaudry, *J. Comput. -Aid. Mol. Des.* **2004**, *18*, 709–718.
[22] P. J. Smith, P. L. A. Popelier, *Org. Biomol. Chem.* **2005**, *3*, 3399–3407.
[23] R. J. Loader, N. Singh, P. J. O'Malley, P. L. A. Popelier, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1249–1254.
[24] N. Singh, R. J. Loader, P. J. O'Malley, P. L. A. Popelier, *J. Phys. Chem. A* **2006**, *110*, 6498–6503.
[25] P. L. A. Popelier, P. J. Smith, *Eur. J. Med. Chem.* **2006**, *41*, 862–873.
[26] K. Roy, P. L. A. Popelier, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2604–2609.
[27] A. Mohajeri, B. Hemmateenejad, A. Mehdipour, R. Miri, *J. Mol. Graphics Mod.* **2008**, *26*, 1057–1065.
[28] A. Mohajeri, M. H. Dinpajooh, *J. Mol. Struct. (THEOCHEM)* **2008**, *855*, 1–5.
[29] S. T. Howard, O. Lamarche, *J. Phys. Org. Chem.* **2003**, *16*, 133–141.
[30] R. F. W. Bader, H. J. T. Preston, *Int. J. Quant. Chem.* **1969**, *3*, 327–347.
[31] *GaussView3.0*, Semichem Inc., Gaussian Inc., Pittsburgh, PA, USA, **2003**.
[32] *GAUSSIAN03*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. J. Montgomery, J. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Pittsburgh, PA, **2003**.
[33] MORPHY98, a program written by P.L.A., *Popelier with a contribution from R.G.A. Bone*, UMIST, Manchester, England, EU, **1998**.
[34] P. L. A. Popelier, *Chem. Phys. Lett.* **1994**, *228*, 160–164.
[35] UMETRICS, *SIMCA-P 10.0, info@umetrics.com; www.umetrics.com, Umeå, Sweden,* (**2002**).
[36] S. Wold, M. Sjostrom, L. Eriksson, *Chemo. Intell. Lab. Sys.* **2001**, *58*, 109–130.
[37] S. Wold, in *Chemometric Methods in Molecular Design*, (Ed.: H. van de Waterbeemd) VCH, Weinheim, **1995**, 195–218.
[38] A. Golbraikh, A. Tropsha, *J. Mol. Graphics Mod.* **2002**, *20*, 269–276.
[39] P. P. Roy, K. Roy, *QSAR Comb. Sci.* **2008**, *27*, 302–313.
[40] P. P. Roy, J. T. Leonard, K. Roy, *Chemom. Intell. Lab. Sys.* **2008**, *90*, 31–42.
[41] K. Roy, J. T. Leonard, *QSAR Comb. Sci.* **2006**, *25*, 235–251.
[42] S. Wold, M. Sjostrom, M. L. Eriksson, in *Encyclopedia of Computational Chemistry*, Vol. 3 (Ed.: P. v. R. Schleyer) Wiley, Chichester, GB, **1998**, 2006–2020.
[43] U. A. Chaudry, P. L. A. Popelier, *J. Org. Chem.* **2004**, *69*, 233–241.
[44] A. A. Amiri, B. Hemmateenejad, A. Safavi, H. Sharghi, A. R. S. Beni, M. Shamsipur, *Anal. Chim. Acta.* **2007**, *605*, 11–19.

Copyright © 2008 John Wiley & Sons, Ltd.