Thermodynamic Descriptors Derived from Density Functional Theory Calculations in Prediction of Aquatic Toxicity

Martin Smieško* and Emilio Benfenati

Department of Environmental Chemistry, Institute for Pharmacological Research "Mario Negri", Via Eritrea 62, 20157 Milano, Italy

Received October 19, 2004

Ab initio Density Functional Theory methods at the B3LYP/6-311+G(2d,p)//HF/3-21G(d) level in the gas and solvent phases (Conductor-like Polarizable Continuum Model) were used to calculate the thermodynamic and charge related descriptors including the first-order dissociation constant for a set of 53 substituted phenols. The absolute values of the $pK_{a\ DFT}$ descriptor showed significant deviations from the experimental values of dissociation constants, but the relative trends were well-described ($R^2 = 0.929$, $n = 40$). The descriptors were used for the design of predictive models for two aquatic species *Pimephales promelas* and *Tetrahymena pyriformis*. In both cases the best performing 2-descriptor models were based on the descriptors ΔH^{298}_{gas} and *ClogP* and showed good predictivity with $R^2 = 0.853$, $R^2_{CV} = 0.834$ for *Pimephales* and $R^2 = 0.864$, $R^2_{CV} = 0.833$ for *Tetrahymena*. The descriptors, models and the mode of action of the compounds are discussed.

INTRODUCTION

McFarland's principle about the activity/toxicity of a compound, which is defined as a sum of a compound's abilities to penetrate and react with biomolecular structures, has been known for more than 30 years.[1] While numerous quantitative structure−activity studies confirm that the first part of the principle − compound's penetration − is usually best described by the logarithm of partition coefficient logP, the reactivity still remains a very complex problem, as the number of structural moieties (e.g. functional groups and larger molecular fragments) capable of interacting with biomolecular structures is theoretically infinite. If the system under the study can be reduced to a set of compounds sharing common structural motif directly linked to the reactivity (e.g. functional group − aldehyde, alcohol, carboxyl), the crucial task remains to find the descriptors with the best descriptive value.

In the past decades the rapidly developing computational systems opened the way for ab initio techniques into the fields such as QSAR, where its application was previously hampered by the unacceptably high computational costs. The descriptors derived from ab initio calculations can be either completely theoretical (e.g. partial atomic charge, orbital energies) or directly comparable to a real physicochemical property, that can be measured experimentally (e.g. dissociation constant). While the former are more frequently used, the later allow validation or scaling to the experimental values. Despite the fact that the main application area of quantum chemistry remains the gas-phase calculations, significant progress has been seen in the development of theory of calculations in the solvated phase, i.e., models capable to describe the behavior of a molecule in the solvent environment. This can be proven by a growing number of publications reporting the results of the high-level calculations in the solvated phase with excellent correlation to the experimental values.[2−6]

The main goal of this study was to assess the performance of the thermodynamic descriptors derived from the theoretical ab initio/Density Functional Theory (DFT) calculations in both the gas and the solvated phases, as the descriptors of reactivity of the hydroxyl group of substituted phenols for the prediction of their aquatic toxicity and modes of action (MOA). Special attention was given to the calculated dissociation constant pK_a as an alternative to experimental values.

COMPUTATIONAL DETAILS

Data Set. A search of the U.S. EPA MED-Duluth Fathead Minnow Database,[7] Ecotox Database[8] and literature[9] was performed in order to identify all substituted phenols. The phenol derivatives containing other known reactive functional groups (e.g. carboxyl acids, aldehydes) were discarded. The acute toxicity LC_{50} for fish *Pimephales promelas* (96-hour flow-through test assessing aqueous concentration leading to the death of 50% of exposed subjects) was used as the endpoint. The search yielded 53 compounds containing at least one phenolic OH group. The compounds included in the final study set exhibit narcotic ($n = 33$), reactive ($n = 5$), mixed ($n = 5$) and respiratory uncoupling ($n = 10$) modes of action. Additionally, the search in the literature showed that for 45 phenols present in the *Pimephales* set also the IGC_{50} toxicity values to ciliate *Tetrahymena pyriformis* (concentration inhibiting the growth of 50% of exposed subjects) and MOA could be found.[9−12] The compounds exhibit narcotic ($n = 23$), respiratory uncoupling ($n = 7$), pro-electrophilic ($n = 4$) and soft electrophilic ($n = 4$) modes of action (MOA was unavailable for $n = 7$ compounds). The toxicity values were expressed as the negative of the

* Corresponding author phone: +39 02 39 014 420; fax: +39 02 39 001 916; e-mail: smiesko@marionegri.it.

10.1021/ci049684a CCC: $30.25
Published on Web 02/02/2005

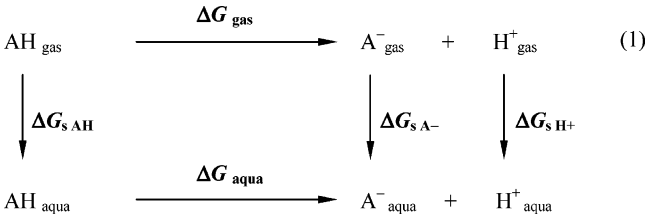

Figure 1. Thermodynamic cycle used for calculation of dissociation constant pK_{aDFT}.

logarithm of LC_{50} for *Pimephales promelas* and IGC_{50} for *Tetrahymena pyriformis* ($-\log LC_{50}$ and $-\log IGC_{50}$, the concentration expressed in mmol.l^{-1}).

Molecular Modeling. The structures of simple substituted phenols were modeled in the ideal planar arrangement (e.g. nitrophenols). Data from crystal structure database (Ref. Id: SAXYEV[13]) were used to construct the starting geometries of more complex compounds containing intramolecular hydrogen bond (2,2′-methylenebis(3,4,6-trichlorophenol) and 2,2′-methylenebis(4-chlorophenol)). In situations where multiple or strained conformations could exist a conformational search was used to construct the optimal starting geometry. If multiple minima were found, only the lowest energy conformation was selected for the calculation of thermodynamic properties.

The geometry of every compound and its respective anion (Ph-OH and Ph-O^-) was completely optimized at the Restricted Hartree−Fock (RHF) level with the simple double-ζ 3-21G(d) basis set.[14] To ensure that the found conformation was a minimum and not a transition state (i.e. the structures represented stable gas-phase minima) at each optimized geometry vibrational analysis was performed at the same level of theory RHF/3-21G(d). Next followed a single point energy calculation at the DFT RB3LYP level with the triple-ζ basis set with diffuse and polarization functions 6-311+G(2d,p) using a tight convergence criteria SCF=Tight.[15−18] This computational procedure was proven to provide good quality results at fair computational costs,[19] thus being suitable for descriptor calculation even for large data sets. To calculate the thermodynamic properties in the solvent environment (water), single point calculation with Conductor-like Polarizable Continuum solvation Model (CPCM)[20−24] as implemented in Gaussian 03W was used in combination with DFT calculation at the B3LYP/6-311+G-(2d,p) level. All geometry optimizations and property calculations were performed by means of Gaussian 03W package of computer codes.[25]

Natural Bond Orbital analysis (keyword: POP=NBO), as implemented in Gaussian 03W − NBO version 3.1,[26] as well as Mulliken analysis were performed at the optimized geometries in order to calculate the charge distribution in the molecules.

Descriptor Calculation. First, the thermodynamic properties needed for calculation of the dissociation constant pK_a were evaluated. These included the gas-phase energies at the DFT level and the solvation energies at the DFT level with the CPCM solvation model. Based on these values the following descriptors were calculated: the dissociation constant at the DFT level with the CPCM solvation model p$K_{a\ DFT}$, the solvation free energies of the neutral molecule, anion and their difference $\Delta G_{s\ AH}$, $\Delta G_{s\ A-}$, and $\Delta\Delta G_s$, and the DFT gas-phase deprotonation enthalpy $\Delta H^{298}{}_{gas}$ at standard temperature $T = 298.15$ K.

The first-order dissociation constant p$K_{a\ DFT}$ values were calculated considering the thermodynamic cycle described in Figure 1, which previously was successfully applied to studying pK_a of substituted phenols[6] and following equations:

$$pK_a = -\log K_a$$

$$\Delta G° = -2.303RT \log K_a$$

$$pK_a = \Delta G°/2.303RT$$

$$\Delta G° = \Delta G_{aqua}$$

$$\Delta G_{aqua} = \Delta G_{gas} + \Delta\Delta G_s$$

where ΔG_{gas} is the free energy of deprotonation in the gas phase and $\Delta\Delta G_s$ is the difference of the solvation energies of the neutral molecule, anion and proton $\Delta\Delta G_s = \Delta G_{s\ AH} - \Delta G_{s\ A-} - \Delta G_{s\ H+}$. The thermodynamic terms for the proton particle were taken from the literature[3,27] ($G_{gas\ H+} = -26.28$ kJ.mol^{-1}, $\Delta G_{s\ H+} = 1107.13$ kJ.mol^{-1}). The enthalpy for the gas-phase deprotonation reaction $\Delta H^{298}{}_{gas}$ at $T = 298.15$ K (see Figure 1) was calculated from the following equations:

$$\Delta H = \Delta E + \Delta pV$$

$$\Delta H = (E_{A-} + 3/2RT) - E_{AH} + RT$$

$$\Delta H = E_{A-} - E_{AH} + 5/2RT$$

where the term ΔpV is replaced by RT, because one mol of gas is obtained in the deprotonation reaction 1 (Figure 1). As the electronic energy of H^+ is zero, it is replaced by its translational energy equal to $3/2RT$.

The descriptors energy of the highest occupied and lowest unoccupied molecular orbital E_{HOMO}, E_{LUMO}, and their difference $\Delta E_{HOMO-LUMO}$, and charges on the oxygen atom and the hydrogen atom of hydroxyl group of phenols calculated by both NBO and Mulliken population analyses $Q_{O\ NBO}$, $Q_{H\ NBO}$, $Q_{O\ Mulliken}$, $Q_{H\ Mulliken}$, were extracted from the output files. In addition to these, the theoretical values of logarithm of partition coefficient were calculated by the *ClogP* algorithm of DayLight.[28]

To compare the performance of the descriptor p$K_{a\ DFT}$, the first-order dissociation constants pK_a were calculated also by the fragment contribution methods implemented in the renowned packages PALLAS 3.0[29] and ACD/Labs version 6.0[30] (p$K_{a\ PALLAS}$, p$K_{a\ ACD/Labs}$). A search in the Syracuse Research Corporation database[31] was performed, yielding values of experimental dissociation constant pK_a for 40 phenols.

In total 31 descriptors were constructed and imported to the QSAR package CODESSA 2.21.[32]

Design of Predictive Models. For both endpoints the performance of the 31 thermodynamic and charge-related descriptors was evaluated with the following procedure: as the importance of the logP descriptor for the QSAR studies is generally known, first a single descriptor model employing *ClogP* was calculated. The *ClogP* was found to be the best correlated descriptor with the toxicity for both endpoints with

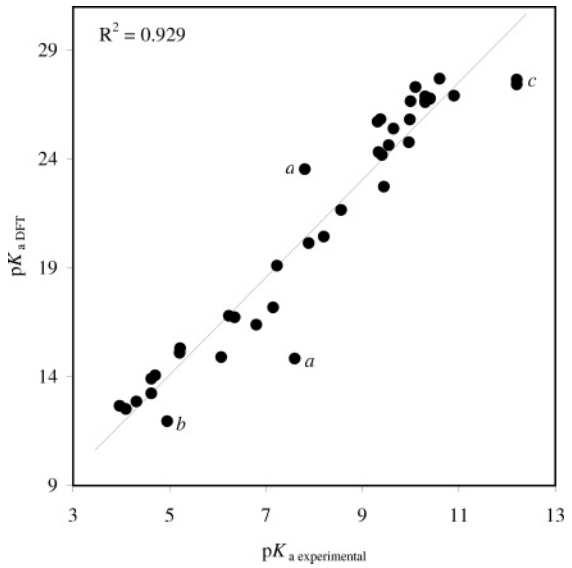

Figure 2. The plot of pK_a experimental vs calculated p$K_{a\ DFT}$.

correlation coefficient $R^2 = 0.789$ for *Pimephales* and $R^2 = 0.680$ for *Tetrahymena*. Then all the remaining descriptors were used one by one to improve the predictivity of the single descriptor model. In this way the 2-descriptor models for all compounds in the set ($n = 53$) were constructed as the best fit against experimental values of toxicity ($-\log LC_{50}$ and $-\log IGC_{50}$, concentration in mmol.l^{-1}). The 2-descriptor models were found to be optimal based on the *F*-test value, because the 3- and 4-descriptor models found by the Best Multiple Linear Regression analysis showed lower *F*-test values (*Pimephales*: 2-descriptor F = 144.6, 3-descriptor F = 117.6, 4-descriptor F = 95.6; *Tetrahymena*: 2-descriptor F = 133.4, 3-descriptor F = 97.5, 4-descriptor F = 80.1). The quality of the models based on 2 descriptors was evaluated by the correlation coefficient R^2 and validated by the leave-1-out method implemented in CODESSA (expressed by the cross-validated correlation coefficient R^2_{CV}).

RESULTS AND DISCUSSION

Dissociation Constant pK_a. The absolute values of the dissociation constant calculated at the DFT level p$K_{a\ DFT}$ differed significantly from the experimental values, which can be expected due to the approximations in the thermodynamic cycle and the level of theory used for the calculation (the calculation of pK_a is very sensitive, as the error of 5.7 kJ·mol^{-1} in $\Delta G°$ value causes the error of 1 pK_a unit). However, the relative trends showed a good correlation to experimental pK_a with the equation $pK_a = 0.414 \cdot pK_{a\ DFT} - 0.607$ ($R^2 = 0.929$, $R^2_{CV} = 0.920$, F = 495.7, $s^2 = 0.421$, $n = 40$). The plot of the experimental vs calculated values is in Figure 2. The highest relative errors were observed with the compounds 2,2′-methylenebis(3,4,6-trichlorophenol) and 2,2′-methylenebis(4-chlorophenol), which were modeled with an intramolecular hydrogen bond based on the crystal structure of a similar compound. This form of internal stabilization seems to be of less importance in the solution, where the conformation can be stabilized by bridging water molecules, which results in the underestimation of dissociation constants of both compounds calculated on the basis of gas-phase conformation. However, the dissociation constants for these two compounds were inaccurately predicted (absolute error ~ 1.5 pK_a unit) also by fragment based approaches of PALLAS and ACD/Labs pK_a software. A large negative error in the calculation of pK_a for 2,6-di-(*tert*)-butyl-4-methylphenol and 2,4,6-tri-(*tert*)-butylphenol apparently stems from the problematic calculation of solvent effects near two bulky alkyl substituents in the *ortho* position to the phenolic hydroxyl group, because *para* (*tert*)-butyl subsubstituted derivatives were predicted more accurately. The large positive deviation of calculated dissociation constant in the case of 4-amino-2-nitrophenol is caused by the presence of a basic amino group at the *para* position.

Table 1. Correlation Coefficients of pK_a Descriptors on a Common Set of 37 Substituted Phenols

species	correlation coefficient R^2			
	p$K_{a\ DFT}$	p$K_{a\ ACD/Labs}$	p$K_{a\ PALLAS}$	p$K_{a\ EXP.}$
Pimephales promelas	0.248	0.097	0.079	0.147
Tetrahymena pyriformis	0.475	0.231	0.212	0.301

Table 2. Statistical Data for the 10 Best Performing Descriptors

Pimephales promelas ($n = 53$)			*Tetrahymena pyriformis* ($n = 45$)		
descriptor	R^2	F value	descriptor	R^2	F value
ClogP	0.789	190.4	ClogP	0.680	91.4
$\Delta G_{s\ anion}$	0.619	82.7	$\Delta G_{s\ anion}$	0.591	62.2
$\Delta\Delta G_s$	0.440	40.0	$\Delta\Delta G_s$	0.563	55.3
ΔH^{298}_{gas}	0.311	23.1	ΔH^{298}_{gas}	0.523	47.1
$\Delta G_{s\ neutral}$	0.216	14.0	p$K_{a\ DFT}$	0.410	29.9
p$K_{a\ DFT}$	0.179	11.1	$E_{HOMO\ anion\ gas}$	0.355	23.7
$E_{HOMO\ anion\ gas}$	0.154	9.3	p$K_{a\ ACD/Labs}$	0.218	12.0
$E_{LUMO\ anion\ gas}$	0.082	4.6	$E_{LUMO\ anion\ gas}$	0.204	11.0
p$K_{a\ ACD/Labs}$	0.069	3.8	p$K_{a\ PALLAS}$	0.196	10.5
$Q_{H\ NBO\ gas}$	0.059	3.2	$Q_{O\ Mulliken\ anion\ CPCM}$	0.182	9.6

To objectively evaluate the performance of the different methods for calculation of dissociation constant a common subset of 37 compounds was prepared, with both experimental toxicity and dissociation constant values available for all compounds. The statistical values of this analysis are summed in Table 1. Surprisingly, the pK_a values obtained from DFT calculation exhibited better correlation to toxicity than the experimental pK_a values, which could be caused by variability of the experimental procedures used for pK_a assessment, as the results in the Syracuse PhysProp database come from different laboratories from the period of the last 4 decades. Values produced by fragment based methods showed the lowest correlation with toxicity.

Descriptors. Table 2 contains the statistical data for the 10 best-performing descriptors. For both biological endpoints the order of the first 4 best-correlated descriptors is the same. Along with *ClogP*, the descriptors solvation free energy of phenoxide anion $\Delta G_{s\ anion}$, difference of solvation free energies $\Delta\Delta G_s$ and gas-phase deprotonation enthalpy ΔH^{298}_{gas} showed high to moderate correlation with the toxicity. While the *ClogP* and ΔG_s descriptors are linked with the entropic effects, i.e., willingness of a molecule to persist in the solvated state in the water phase, the descriptors pK_a, $\Delta\Delta G_s$ and especially ΔH^{298}_{gas} (descriptors p$K_{a\ DFT}$ and $\Delta\Delta G_s$ are significantly correlated with the descriptor ΔH^{298}_{gas} with $R^2 = 0.914$ and $R^2 = 0.831$) describe the thermodynamic properties related to the compound's reactivity in terms of

its ability to release the proton from the hydroxyl group and form a phenoxide anion.

The descriptors E_{HOMO} and E_{LUMO} are related to the compound's nucleophilic and electrophilic reactivity, and the charge descriptors of the oxygen and hydrogen atoms Q_O and Q_H reflect the overall effect of the electron withdrawing/donating substituents to the charge distribution on the hydroxyl or phenoxide functional groups.

Interestingly, some molecular descriptors (e.g. ΔG_s, E_{HOMO}, E_{LUMO}) showed much better correlation with the toxicity, if calculated for the deprotonated form of studied compounds. The most significant differences were observed in case of the descriptor the free energy of solvation ΔG_s: its correlation with the toxicity was $R^2 = 0.619$ when calculated for a deprotonated form of phenols vs $R^2 = 0.216$ for a neutral form for *Pimephales* and $R^2 = 0.591$ vs $R^2 = 0.057$ for *Tetrahymena*. This rules out a chance correlation but rather supports the hypothesis that for a toxic action of phenols ionic (deprotonated) forms are important and should be considered in the development of the QSAR models. Indeed, at physiological p*H* chlorophenols and nitrophenols with $pK_a < 7$ are prevalently present in the ionized form, which is likely the bioactive form, i.e., the form responsible for their toxic effect.

Models. Table 3 contains the statistical data for the first 10 best predictive 2-descriptor models. The results showed that for both biological endpoints the best 2-descriptor models were obtained by combining the *ClogP* descriptor, which is linked to a compound's penetration abilities, with the descriptor gas-phase deprotonation enthalpy ΔH^{298}_{gas} linked to the reactivity of the compound's hydroxyl group. This observation is in good agreement with a previous study, in which the ΔH_{gas} descriptor calculated by the semiempirical quantum chemical method AM1 was applied along with 3 other thermodynamic descriptors to study the toxicity of 8 phenols (phenol, 5 chlorophenols, 2 nitrophenols) and proved a significant contribution to toxicity variation with 10 biological test systems.[33]

A combination of *ClogP* with the descriptors $pK_{a\ DFT}$ or $\Delta\Delta G_s$ yielded models with slightly lower correlation coefficients R^2 when compared to a model based on *ClogP* with ΔH^{298}_{gas}. Both descriptors $pK_{a\ DFT}$ and $\Delta\Delta G_s$ include the solvation effects (the entropic contribution from the ΔG_s terms), which duplicate the information of the descriptor *ClogP* (*logP* can be decomposed to ΔG_s and the *Contact Surface Area* terms[33]), while the information value of the descriptor ΔH^{298}_{gas} is higher, as it describes the electronic effects of the substituents on the deprotonation of the hydroxyl group in the gas phase, i.e., unperturbed by the solvent environment. The dissociation constants $pK_{a\ PALLAS}$ and $pK_{a\ ACD/Labs}$ calculated by the fragment-based methods implemented in PALLAS and ACD/Labs software packages are very similar (intercorrelation $R^2 = 0.993$), but neither of the two provided better predictive model in combination with the *ClogP*, when compared to the results obtained with the descriptor $pK_{a\ DFT}$ along with *ClogP*. Nevertheless, the dissociation constant pK_a in combination with *ClogP* proved to be a valuable descriptor of reactivity suitable for the design of predictive models in toxicology, as suggested by previous studies on phenols for *Tetrahymena*.[34,35]

The "traditional" descriptors of molecular reactivity the energy of the frontier molecular orbitals E_{HOMO} and E_{LUMO} and the charge derived descriptors Q_O and Q_H, whose suitability for the design of predictive models and discrimination of mode of action has already been confirmed several times in the past in various fields of QSAR,[10,36−40] in this study did not provide better results than the thermodynamic descriptors ΔH^{298}_{gas} and $pK_{a\ DFT}$. However, the descriptor energy of the highest occupied molecular orbital for phenoxide anion calculated in the gas phase $E_{HOMO\ anion\ gas}$ and solvated phase with the CPCM model $E_{HOMO\ anion\ CPCM}$ showed its good predictive power especially in models for *Tetrahymena*, because for some substituted phenols the electrophilic properties, which are well-described by the molecular orbital descriptors, are important for their toxic action.[36−40]

The descriptor $\Delta G_{s\ anion}$ alone is well-correlated with toxicity, which was proven also in a study on chlorophenols using semiempirical quantum chemical methods, because the descriptor free energy of solvation ΔG_s was found to be directly linked to toxicity for 5 biological endpoints.[41] Nevertheless, its combination with the descriptor *ClogP* in a predictive model is not feasible, as the two descriptors are considerably intercorrelated ($R^2 = 0.6$) and the resulting model shows only minor improvement of predictivity (from $R^2 = 0.789$ to $R^2 = 0.813$).

The final QSAR equations and the statistical parameters of the best 2-descriptor models for the two biological endpoints are as follows:

$$-\log LC_{50}\ (\textit{Pimephales}) = 0.554 \cdot \textit{ClogP} - 0.00567 \cdot \Delta H^{298}_{gas} + 8.02$$

$$n = 53,\ R^2 = 0.853,\ R^2_{CV} = 0.834,\ F = 144.6,\ s^2 = 0.174$$

$$-\log IGC_{50}\ (\textit{Tetrahymena}) = 0.423 \cdot \textit{ClogP} - 0.00884 \cdot \Delta H^{298}_{gas} + 12.5$$

$$n = 45,\ R^2 = 0.864,\ R^2_{CV} = 0.833,\ F = 133.4,\ s^2 = 0.135$$

The predicted toxicity values for these two equations along with the experimental values are listed in the Table 4. The best performing model for each species was validated by a more rigorous procedure than leave-one-out, when the original data set was ordered by increasing toxicity and divided into 3 subsets (A, B, and C). For subset A, every third compound in the ordered set starting from the first compound was selected. Similarly, for the subsets B and C every third compound in the ordered set starting from the second and the third compound was selected. Each subset was then used for validation of the model re-trained on the remaining 2/3 of the set (e.g. when subsets A+B were used for training, subset C was used for validation). By this procedure, every compound was twice included in the training set and once in the validation set. The correlation coefficients obtained on the validation subsets $R^2_{L1/3} = 0.841$ (*Pimephales*) and $R^2_{L1/3} = 0.842$ (*Tetrahymena*) confirmed very good predictivity of the two best-performing QSAR models.

Descriptors and the Mode of Action. Figures 3 and 4 show the plots of the *ClogP* descriptor vs ΔH^{298}_{gas} descriptor of the best performing 2-descriptor models for both biological

Table 3. Statistical Data for the 10 Best 2-Descriptor Models

Pimephales promelas ($n = 53$)				*Tetrahymena pyriformis* ($n = 45$)			
ClogP + descriptor	R^2	F value	R^2_{CV}	*ClogP* + descriptor	R^2	F value	R^2_{CV}
ΔH^{298}_{gas}	0.853	144.6	0.834	ΔH^{298}_{gas}	0.864	133.4	0.833
$\Delta\Delta G_s$	0.852	144.0	0.830	$pK_{a\ DFT}$	0.853	122.0	0.827
$pK_{a\ DFT}$	0.841	132.2	0.822	$E_{HOMO\ anion\ gas}$	0.850	118.7	0.815
$Q_{H\ NBO\ gas}$	0.835	126.3	0.813	$E_{HOMO\ anion\ CPCM}$	0.843	112.6	0.815
$E_{HOMO\ anion\ gas}$	0.833	124.4	0.813	$\Delta\Delta G_s$	0.834	105.6	0.796
$pK_{a\ ACD/Labs}$	0.830	121.8	0.809	$E_{LUMO\ neutral\ gas}$	0.827	100.6	0.800
$E_{HOMO\ anion\ CPCM}$	0.830	121.8	0.809	$E_{LUMO\ anion\ CPCM}$	0.826	99.7	0.799
$pK_{a\ PALLAS}$	0.829	121.1	0.808	$E_{LUMO\ neutral\ CPCM}$	0.823	97.6	0.796
$Q_{H\ Mulliken\ gas}$	0.827	119.3	0.805	$Q_{O\ NBO\ anion\ CPCM}$	0.815	92.7	0.785
$E_{LUMO\ neutral\ gas}$	0.821	114.5	0.796	$pK_{a\ ACD/Labs}$	0.815	92.3	0.786

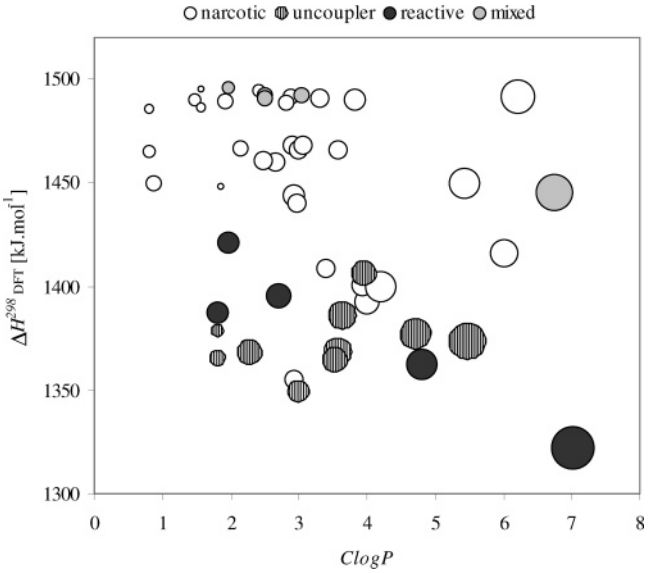

Figure 3. The plot of the *ClogP* vs ΔH^{298}_{DFT} descriptor indicating the mode of action for *Pimephales promelas*. The size of the circle is proportional to the toxicity of the compound. The bigger the circle, the more toxic the compound.

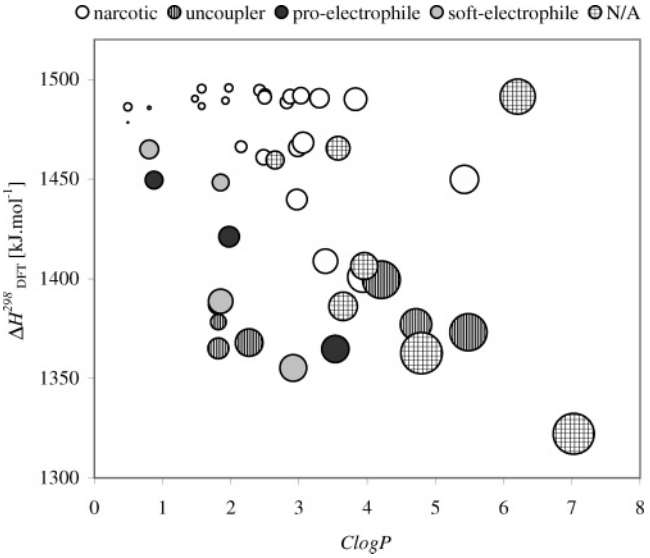

Figure 4. The plot of the *ClogP* vs ΔH^{298}_{DFT} descriptor indicating the mode of action for *Tetrahymena pyriformis*. The size of the circle is proportional to the toxicity of the compound. The bigger the circle, the more toxic the compound.

species indicating the toxicity and the MOA, which has been assigned to the compounds in previous SAR studies.[7,9−14] However, some compounds have borderline values of physicochemical descriptors, which means that their actual toxic effect is more likely a sum of partial toxic effects of different nature than a single MOA.

The narcotic mode of action is frequently observed with the compounds with high values of ΔH^{298}_{gas} descriptor in the range between 1388.6 and 1495.3 kJ.mol^{-1} for *Pimephales* and between 1400.8 and 1495.7 kJ.mol^{-1} for *Tetrahymena*. This means that the deprotonation of the hydroxyl groups of these compounds is not favorable, which facilitates the passive transport of the uncharged form through the membranes to the central nervous system and lipid tissues with a rate depending mainly on the value of the partition coefficient (a single exception to the rule for *Pimephales* is 3-trifluoromethyl-4-nitrophenol with high acidity caused by trifluoromethyl substituent; $\Delta H^{298}_{gas} = 1355.1$ kJ.mol^{-1}). Similarly, these compounds have higher experimental values of dissociation constant pK_a (mean value 8.9 units) implying that at physiological p*H* the balance is shifted toward the neutral form. The toxicity of this group of compounds can be well-predicted by the *ClogP* descriptor alone (*Pimephales*: $R^2 = 0.821$; *Tetrahymena*: $R^2 = 0.815$), with values ranging from 0.5 to 6.2 units for *Pimephales* and from 0.5 to 5.4 for *Tetrahymena*. This finding is in good agreement with a study performed on 60 phenols for *Tetrahymena*, when a highly predictive model based only on the logarithm of K_{OW} could be designed for a subset of 17 phenols with $pK_a > 9.8$ ($R^2 = 0.97$, $F = 487.6$).[42]

The compounds with the reactive (*Pimephales*), electrophilic and pro-electrophilic (*Tetrahymena*) modes of action usually exhibit lower values of the ΔH^{298}_{gas} descriptor when compared to the compounds of similar values of *ClogP* with the narcotic mode of action, which means that these compounds can be more easily deprotonated to form the reactive phenoxide anions, which on the other hand cannot penetrate passively through the biomembranes.

The compounds with the respiratory uncoupling MOA, which impair the proton gradient on the mitochondrial membrane, are by definition lipophilic weak acids. In the present study these compounds were found to have low calculated values of the ΔH^{298}_{gas} descriptor when compared to other groups, indicating their weak acidic character (mean value of experimental dissociation constant pK_a exp. is 4.9 units for *Pimephales* and 4.8 for *Tetrahymena*). The gas-phase acidity of the phenolic hydroxyl group seems to determine the MOA of this group, because for all uncoupling compounds the ΔH^{298}_{gas} values lie in a relatively narrow range between 1349.1 and 1406.3 kJ.mol^{-1} for *Pimephales* (33.0% of the total range) and between 1365.0 and 1399.4 kJ.mol^{-1} for *Tetrahymena* (19.8% of the total range), while

Table 4. Experimental and Predicted Toxicity Values for 53 Substituted Phenols

	toxicity [mmol.L^{-1}]			
compound	exptl −log LC_{50}	pred −log LC_{50}	exptl −log IGC_{50}	pred −log IGC_{50}
2,4-dinitrophenol	1.14	1.29	1.06	1.22
2,3,4,6-tetrachlorophenol	2.35	2.35	N/A	N/A
4-chloro-3-methylphenol	1.29	1.37	0.80	0.82
2,2′-methylenebis(3,4,6-trichlorophenol)	4.29	4.42	3.04	3.80
4,4′-isopropylidenebis(2,6-dichlorophenol)	2.44	3.33	N/A	N/A
4-(*tert*)-pentylphenol	1.80	1.70	1.23	0.97
pentachlorophenol	3.08	2.83	2.07	2.34
2,4,6-trichlorophenol	1.33	1.92	1.41	1.50
3-trifluoromethyl-4-nitrophenol	1.36	1.96	1.65	1.77
2-nitrophenol	−0.06	0.84	0.67	0.50
2-(*sec*)-butyl-4,6-dinitrophenol	2.54	2.25	N/A	N/A
1-naphthol	1.49	1.22	0.75	0.73
2-phenylphenol	1.44	1.40	1.09	0.83
2-methylphenol	0.89	0.65	−0.29	0.16
2-chlorophenol	0.97	0.91	0.18	0.47
2,2′-methylenebis(4-chlorophenol)	2.94	2.96	3.09	2.50
4-(*tert*)-butylphenol	1.46	1.40	0.91	0.74
4-nitrophenol	0.38	1.18	1.42	1.02
4-nonylphenol	3.20	3.01	2.47	1.96
2,4-dimethylphenol	0.87	0.89	0.14	0.33
4-methylphenol	0.82	0.64	−0.16	0.13
4-chlorophenol	1.32	1.12	0.54	0.65
phenol	0.51	0.39	−0.32	−0.03
2,4,6-tribromophenol	1.70	2.27	2.03	1.80
4-amino-2-nitrophenol	0.63	0.17	0.88	−0.09
catechol	1.08	0.29	0.75	0.07
2,4-dichlorophenol	1.32	1.51	1.04	1.05
4-ethylphenol	1.07	0.95	0.21	0.38
2,6-di-(*tert*)-butyl-4-methylphenol	2.78	2.81	1.79	2.00
3-methoxyphenol	0.22	0.47	−0.33	0.04
4-methoxyphenol	0.05	0.42	−0.14	−0.04
2,5-dinitrophenol	1.74	1.17	0.95	1.02
2,4,6-trimethylphenol	1.02	1.16	0.42	0.55
4,6-dinitro-2-methylphenol	2.01	1.53	1.73	1.38
2,6-dinitrophenol	0.67	1.22	0.54	1.10
pentabromophenol	3.72	3.28	2.66	2.70
4-propylphenol	1.09	1.25	0.64	0.61
2,4,6-tri-(*tert*)-butylphenol	3.63	3.57	N/A	N/A
4-phenoxyphenol	1.58	1.70	1.36	1.07
tetrachlorocatechol	2.29	2.25	1.70	1.95
4-phenylazophenol	2.23	2.24	1.66	1.76
2-allylphenol	0.95	0.96	0.33	0.39
4-chlorocatechol	1.96	1.06	1.06	0.79
2,3,6-trimethylphenol	1.22	1.15	0.28	0.55
4,5-dichloro-2-methoxyphenol	1.64	1.46	N/A	N/A
4,5-dichlorocatechol	2.30	1.61	N/A	N/A
2,3,4,5-tetrachlorophenol	2.75	2.42	2.72	1.93
2,4-dinitro-1-naphthol	1.84	2.03	N/A	N/A
3-acetamidophenol	−0.87	−0.13	−0.16	−0.41
4-acetamidophenol	−0.73	−0.08	−0.82	−0.34
4-(*tert*)-butyl-2,6-dinitrophenol	2.65	2.19	1.80	1.81
1,3-benzenediol	0.34	0.05	−0.65	−0.28
4,4′-dihydroxydiphenyl ether	1.54	1.31	N/A	N/A

ClogP values largely varied from 1.8 to 5.5 units with both species (56.1% of the total range).

In general, the compounds causing narcosis are significantly less toxic, when compared to the reactive or uncoupling ones. The compounds, which have either very high lipophilicity (increased penetration, e.g. 4-nonylphenol and 2,4,6-tri-(*tert*)-butylphenol) or pronounced gas phase acidity (increased reactivity, e.g. 4,6-dinitro-2-cresol and 3-trifluoromethyl-4-nitrophenol) or a combination of these two physicochemical properties are highly toxic (pentabromophenol and 2,2′-methylenebis(3,4,6-trichlorophenol)) to both aquatic species.

CONCLUSION

In the data set of 53 phenols, the absolute values of the dissociation constant pK_a calculated at the B3LYP/6-311+G-(2d,p)//RHF/3-21G* level of theory with the CPCM solvation model were found to be significantly deviated from the values assessed by the experimental procedure; however, the relative trends were well reproduced, with correlation coefficient $R^2 = 0.929$. The relative error stems from the approximations in the thermodynamic cycle and the level of theory used for the property calculation. In a few cases significant relative deviations were caused by particular structural properties of the studied compounds (steric effects, hydrogen bonding). In general ab initio/DFT calculated pK_a values showed superiority over fragment-based and even experimental data in terms of the correlation to toxicity and performance of QSAR models.

In the set of 31 ab initio/DFT calculated thermodynamic descriptors related to the deprotonation of phenolic compounds the descriptor gas-phase deprotonation enthalpy of the hydroxyl group ΔH^{298}_{gas}, which could be regarded as a descriptor of reactivity, in combination with the descriptor *ClogP* describing compound's penetration abilities, showed very good performance in toxicity prediction for two different aquatic species ($R^2 > 0.85$, $R^2_{CV} > 0.83$, $R^2_{L1/3O} > 0.84$). The QSAR models based on the 2 descriptors could be clearly explained with respect to the mode of action of this group of chemicals.

The level of ab initio/DFT theory used in this study was selected with respect to computational equipment and feasibility of the study. The room for the improvement still exists, especially if high-level theoretical methods would be applied in the calculation of thermodynamic properties of the studied molecules. However, the QSAR models demonstrated that a good relative estimation of the deprotonation thermodynamics is more important than the absolute accuracy of calculated physicochemical properties.

The results indicate that for compounds with the acidic character the ab initio/DFT thermodynamic descriptors, based on both gas and solvated phase calculations, linked to the proton dissociation of the acidic functional group can be suitable for the design of the predictive QSAR models in environmental toxicology of aquatic species.

ACKNOWLEDGMENT

Support from European Union within the DEMETRA project is gratefully acknowledged. Authors thank Prof. Alan Katritzky (University of Florida, Gainesville, USA) and Prof. Mati Karelson (University of Tartu, Estonia) for providing CODESSA software.

REFERENCES AND NOTES

(1) McFarland, J. W. On the Parabolic Relationship Between Drug Potency and Hydrophobicity. *J. Med. Chem.* **1970**, *13*, 1192−1196.

(2) Jang, Y. H.; Sowers, L. C.; Cagin, T.; Goddard, W. A., III. First Principles Calculation of pK_a Values for 5-Substituted Uracils. *J. Phys. Chem. A* **2001**, *105*, 274−280.

(3) Liptak, M. D.; Shields, G. C. Accurate pK_a Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined

with CPCM Continuum Solvation Methods. *J. Am. Chem. Soc.* **2001**, *123*, 7314−7319.
(4) Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J. Prediction of the pK_a of Carboxylic Acids using the ab Initio Continuum-Solvation Model PCM-UAHF. *J. Phys. Chem. A* **1998**, *102*, 6706−6712.
(5) Schüürman, G. Quantum chemical analysis of the energy of proton transfer from phenol and chlorophenols to H_2O in the gas phase and in aqueous solution. *J. Chem. Phys.* **1998**, *109*, 9523−9528.
(6) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. Absolute pK_a Determinations for Substituted Phenols. *J. Am. Chem. Soc.* **2002**, *124*, 6421−6427.
(7) Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting Modes of Toxic Action from Chemical Structure: Acute Toxicity in the Fathead Minnow (*Pimephales Promelas*). *Environ. Toxicol. Chem.* **1997**, *16*, 948−967. Web site http://www.epa.gov/med/databases/fathead_minnow.htm
(8) Web site http://www.epa.gov/ecotox/
(9) Ren, S.; Frymier, P. D.; Schultz, T. W. An exploratory study of the use of multivariate techniques to determine mechanisms of toxic action. *Ecotox. Environ. Safe.* **2003**, *55*, 86−97.
(10) Ren, S. Ecotoxicity prediction using mechanism- and nonmechanism-based QSARs: a preliminary study. *Chemosphere* **2003**, *53*, 1053−1065.
(11) Aptula, A. O.; Netzeva, T. I.; Valkova, I. V.; Cronin, M. T. D.; Schultz, T. W.; Kuehne, R.; Schüürmann, G. Multivariate discrimination between modes of toxic action of phenols. *Quant. Struct.-Act. Relat.* **2002**, *21*, 12−22.
(12) Schultz, T. W. TETRATOX: *Tetrahymena* population growth impairment endpoint-a surrogate for fish lethality. *Toxicol. Methods* **1997**, *7*, 289−309.
(13) Ferguson, G.; McCrindle, R.; McAlees, A. J. The synthesis, separation, and crystal and molecular structures of the three bisphenols, an ethanol solvate of one of them, and of a minor byproduct, 6, 6′, 7, 7′-tetrachloro-8, 8′-methylenebis(4H-benzo-1,3-dioxin), prepared by condensation of formaldehyde with 3,4-dichlorophenol. *Can. J. Chem.* **1989**, *67*, 779−785.
(14) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. Ab initio Molecular Orbital Theory. John Wiley: New York, 1986.
(15) Parr, R. G.; Yang, W. Density-functional theory of atoms and molecules. Oxford University Press: Oxford, 1989.
(16) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(17) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(18) Lee, C.; Yang W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(19) Foresman, J. B.; Frisch, Æ. Exploring Chemistry with Electronic Structure Methods. Second Edition. Gaussian, Inc.: Pittsburgh, PA, 1996.
(20) Cammi, R.; Mennucci, B.; Tomasi, J. Fast evaluation of geometries and properties of excited molecules in solution: a Tamm Dancoff Model with application to 4-dimethylbenzonitrile. *J. Phys. Chem. A* **2000**, *104*, 5631−5637.
(21) Cammi, R.; Mennucci, B.; Tomasi, J. Second-order Moller−Plesset analytical derivatives for the Polarizable Continuum Model using the relaxed density approach. *J. Phys. Chem. A* **1999**, *103*, 9100−9108.
(22) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Polarizable dielectric model of solvation with inclusion of charge penetration effects. *J. Chem. Phys.* **2001**, *114*, 5691−5701.
(23) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. Recent developments in the Polarisable Continuum Model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002**, *117*, 43−54.
(24) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures and electronic properties of molecules in solution with the C−PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−681.
(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr.; J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. 2003 Gaussian 03, Revision B.03, Gaussian, Inc.: Pittsburgh, PA.
(26) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold F. NBO Version 3.1.
(27) McQuarrie, D. M. Statistical Mechanics. Harper and Row: New York, 1970.
(28) Daylight Chemical Information Systems, Inc., Mission Viejo, CA. Web site http://www.daylight.com/cgi-bin/contrib/pcmodels.cgi
(29) ComGenex, Budapest, Hungary. Web site http://www.comgenex.com/
(30) ACD/Labs pK_a module version 6.0, Toronto, Ontario, Canada. Web site http://www.acdlabs.com/
(31) PhysProp database, Syracuse Research Corporation, Syracuse, NY. Web site http://www.syrres.com/esc/physdemo.htm
(32) SemiChem Inc.; Shawnee Mission, KS. Web site: http://www.semichem.com/
(33) Schüürmann, G.; Segner, H.; Jung, K. Multivariate mode-of-action analysis of acute toxicity of phenols. *Aquat. Toxicol.* **1997**, *38*, 277−296.
(34) Baláž, Š.; Piršelová, K.; Schultz, T. W. Subcellular pharmacokinetics and its potential for library focusing. *J. Mol. Graph. Model.* **2002**, *20*, 479−490.
(35) Baláž, Š.; Piršelová, K.; Schultz, T. W.; Hermens J. Kinetics of Subcellular Distribution of Multiply Ionizable Compounds: A Mathematical Description and its Use in QSAR. *J. Theor. Biol.* **1996**, *178*, 7−16.
(36) Cronin, M. T. D.; Schultz, T. W. Structure-toxicity Relationships for Phenols to *Tetrahymena pyriformis*. *Chemosphere* **1996**, *32*, 1453−1468.
(37) Cronin, M. T.; Aptula, A. O.; Duffy, J. C.; Netzeva, T. I.; Rowe, P. H.; Valkova, I. V.; Schultz, T. W. Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to *Tetrahymena pyriformis*. *Chemosphere* **2002**, *49*, 1201−1221.
(38) Ren, S.; Schultz, T. W. Identifying the mechanism of aquatic toxicity of selected compounds by hydrophobicity and electrophilicity descriptors. *Toxicol. Lett.* **2002**, *129*, 151−160.
(39) Bearden, A. P.; Schultz, T. W. Structure−activity relationships for Pimephales and Tetrahymena: a mechanism of action approach. *Environ. Toxicol. Chem.* **1997**, *16*, 1311−1317.
(40) Schüürmann, G.; Aptula, A. O.; Kühne, R.; Ebert, R.-U. Stepwise Discrimination between Four Modes of Toxic Action of Phenols in the Tetrahymena pyriformis Assay. *Chem. Res. Toxicol.* **2003**, *16*, 974−987.
(41) Bureau, R.; Faucon, J. C.; Faisant, J.; Briens, F.; Rault, S. Applicability of the free energies of solvation for the prediction of ecotoxicity: study of chlorophenols. *SAR QSAR Environ. Res.* **1997**, *6*, 163−181.
(42) Schultz, T. W.; Bearden, A. P.; Jaworska, J. S. A novel QSAR approach for estimating toxicity of phenols. *SAR QSAR Environ. Res.* **1996**, *5*, 99−112.

CI049684A