New Density Functional and Atoms in Molecules Method of Computing Relative pK_a Values in Solution

Kenneth R. Adam*

School of Pharmacy and Molecular Sciences, James Cook University, Townsville, Australia 4811

Received: July 22, 2002; In Final Form: October 9, 2002

A theoretical structure−property relation between pK_a and Bader's atoms in molecules (AIM) energy of the dissociating proton was obtained by an approximation of the standard gas-phase expression for the equilibrium constant expressed in terms of molecular partition functions. This relation was then tested by solvated density functional computations on a series of aliphatic carboxylic acids, substituted benzoic acids, phenols, anilinium ions, and pyridinium ions using the COSMO solvation model. Comparison with accurate experimental values indicates that average unsigned errors of generally less than 0.2 pK_a units can be achieved in the calculation of relative pK_a values. The inclusion of specifically hydrogen-bonded water molecules in the vicinity of the dissociating proton was found to improve the agreement between theory and experiment greatly. Computed pK_a values for some diprotic acids were also investigated.

1. Introduction

Knowledge of the acid dissociation constants of the ionizable protons in molecules is of fundamental importance in many areas of chemistry and biochemistry as it allows the protonation states of acids to be determined at any particular pH value. Consequently, much effort has been devoted to the experimental determination of pK_a values, and although in many cases accurate experimental measurements can be easily made, there are other situations where accurate measurements are difficult. Hence, there is much interest in developing methodology for predicting pK_a values in a variety of chemical systems by various quantum chemical techniques. In addition, theoretical methods are able to elucidate some of the important factors involved in the relationship between molecular structure and pK_a values.

The dissociation of an acid in aqueous solution may be represented in the form

$$HA + H_2O \rightleftharpoons H_3O^+ + A^- \qquad (1)$$

Here it is understood that H_3O^+ represents a proton together with the associated solvation shell. Combining the definition of pK_a,

$$pK_a = -\log K_a$$

with the standard thermodynamic relation

$$\Delta G^\circ = -RT \ln K_a$$

gives

$$pK_a = \Delta G^0/2.302RT$$

One approach to computing either absolute or relative pK_a values is based upon ab initio quantum chemical calculations of the free energy change ΔG^0 of this process either in a gas or solution. This methodology is well established for the gas phase[1] and in the special case of an ideal gas ΔG^0_{gas} can be computed by a gas-phase geometry optimization followed by a vibrational analysis for each of the species involved in equilibrium 1. This provides the structural data required for the computation of ΔG^0_{gas} using standard expressions for the molecular partition functions of an ideal gas.[2] To compute pK_a values in solution, the procedure is extended by using appropriate thermodynamic cycles in which the computed gas-phase ΔG^0_{gas} values are used together with computed values of the free energies of solvation for each of the species involved in the equilibrium to give ΔG^0_{soln}, the total free-energy change in solution.[3−13] Methods of including solvation effects have been recently reviewed by Cramer and Truhlar.[14] Since an error of only 5.7 kJ mol^{-1} in the value of ΔG^0_{soln} produces an error of 1 pK_a unit, these calculations need to be performed at a high level of theory and hence are computationally expensive, particularly for large molecules. Nevertheless, calculated values of pK_a with errors of less than 0.5 pK_a units have been obtained by this method for several carboxylic acids[9] and phenols.[11]

A second approach[15−23] to the computation of pK_a values is to seek a quantitative structure−property relation that is an empirical relationship, usually in the form of a linear regression, between a chemical or biological property of interest (pK_a in this case) and various structural properties of molecules. Values for the structural properties or descriptors may be obtained either from experiment or from quantum mechanical calculations involving the HA species only. This requires much less computational effort than does the first ab initio approach and hence can be more easily used to treat larger, more complex systems.

The work reported in this paper is in the spirit of this second approach. However, instead of using a completely empirical structure−property relation, a theoretically derived relationship between pK_a and a structural property is obtained by introducing approximations into the standard expression for pK_a expressed in terms of the gas-phase molecular partition functions used in the first ab initio method. This relation is then tested by performing calculations of the relative pK_a values of sets of aliphatic carboxylic acids, substituted benzoic acids, phenols, anilinium ions, and pyridinium ions.

* E-mail: ken.adam@jcu.edu.au.

10.1021/jp026577f CCC: $22.00
Published on Web 11/07/2002

2. Theoretical Basis

For an ideal gas system containing molecules behaving as a system of independent particles with no intermolecular interactions, an expression for the equilibrium constant in terms of molecular partition functions can be obtained from standard textbooks.[2] For the general gas-phase equilibrium

$$0 = \sum_{J=1}^{n} \nu_J A_J$$

involving n species A_J with stoichiometric coefficients ν_J, the equilibrium constant K is given by

$$\ln K = \frac{-\Delta E_0}{RT} + \ln \prod_J \left(\frac{q_J^o}{N_A}\right)^{\nu_J} \tag{2}$$

where R is the gas constant, T is the absolute temperature, N_A is Avogadro's number, and q_J^o is the standard molar partition function of species A_J.

$$\Delta E_0 = \sum_J \nu_J U_J^o$$

where U_J^o is the molar energy difference between the lowest vibrational energy level of species A_J and the state where it has been completely dissociated into its component atoms. In the special case of an ideal gas, relatively simple expressions may be obtained for the molecular partition functions.[2]

A corresponding treatment for solution equilibria would be very complex, and so here a more approximate treatment will be followed. A widely used method for treating the effects of solvation in quantum chemistry is the polarized continuum method (PCM)[24] in which the solvent is treated as a homogeneous continuum polarized by the solute placed in a solvent cavity. The electrostatic interaction between solvent and solute is taken into account in an average manner by including additional terms describing the reaction potential due to the solvent in the one-electron Hamiltonian of the solute. In this sense, this is also a pseudo-independent particle model with no direct solute−solute interactions and the interaction of the solute with the solvent being treated in an average manner. Hence, as a first approximation, it will be assumed that an equation of the same general form as eq 2 may also be applied to solution equilibria; however, in this case, no simple expressions are available for the solution-phase partition functions, and in order to develop this approach further, some reasonable approximations will be made. Note that this lack of simple explicit expressions for the solution-phase partition functions makes it difficult to justify these approximations directly, and in this work, they will be justified only a posteri by the ability (or inability) of the final expression to predict experimental results correctly.

For the particular case of equilibrium 1, the general form of eq 2 becomes

$$\ln K_a = \frac{(U^0_{HA} + U^0_{H_2O} - U^0_{A^-} - U^0_{H_3O^+})}{RT} + \ln\left(\frac{q^0_{H_3O^+}}{q^0_{HA}}\frac{q^0_{A^-}}{q^0_{H_2O}}\right)$$

Regrouping the terms in this equation and using the definition of pK_a gives

$$pK_a = \frac{U^0_{A^-} - U^0_{HA}}{2.303RT} - \log\left(\frac{q^0_{A^-}}{q^0_{HA}}\right) + \frac{U^0_{H_3O^+} - U^0_{H_2O}}{2.303RT} - \log\left(\frac{q^0_{H_3O^+}}{q^0_{H_2O}}\right) \tag{3}$$

The last two terms of eq 3 involve only the species H_3O^+ and H_2O and hence will be constant when the equation is used for a fixed temperature comparison of the pK_a values of a series of acids. Also, since the species HA and A^- differ only by an additional hydrogen atom in HA, their partition functions can be expected to have very similar values, giving a ratio close to unity in the second term. This second term will then be close to zero and will be neglected, this approximation becoming more accurate for large molecules. Equation 3 then reduces to

$$pK_a \approx \frac{U^0_{A^-} - U^0_{HA}}{2.303RT} + C' \tag{4}$$

where C' is a constant at any particular temperature. This constitutes the second major approximation. It is convenient to express this equation in terms of the molar energies E_J obtained from standard quantum chemical codes, that is, the value of the energy per molecule as produced by the code multiplied by Avogadro's number. This molar energy E_J is the energy difference between the minimum in the potential surface and the state in which the molecule has been completely decomposed into nuclei and electrons. The relation between E_J and U_J^o for species A_J is given by

$$E_J = U_J^0 - E_J^{ZPE} + \sum_{i=1}^{n_J} E_i$$

where n_J is the number of atoms in A_J, E_i is the molar quantum chemical energy of atom i, and E_J^{ZPE} is the molar zero-point vibrational energy of A_J, that is, the molar vibrational energy of A_J when in the lowest vibrational state. Note that this expression uses the usual sign conventions whereby E_J, U_J^o, and E_i are negative quantities whereas E_J^{ZPE} measured relative to the minimum of the potential surface is positive. Then

$$U^0_{A^-} - U^0_{HA} = E_{A^-} - E_{HA} + E_H + E^{ZPE}_{A^-} - E^{ZPE}_{HA}$$

The difference between the two zero-point energies should be small and will be assumed to be approximately constant for a series of acids of similar structures, which is a third major approximation. Together with the constant energy of the free hydrogen atom, E_H, these terms may then be incorporated into the constant term of eq 4 to give the expression

$$pK_a \approx \frac{E_{A^-} - E_{HA}}{2.303RT} + C$$

where C is constant at any particular temperature. This equation may be developed further using Bader's theory of atoms in molecules (AIM).[25] This theory provides a method of partitioning a molecule into atomic basins whose properties, including energy, can be calculated using quantum mechanics. Furthermore, according to this theory, any molecular property is the sum of the values of the property for the individual partitioned

atoms, and it has been demonstrated that the properties of the atoms defined in this way are roughly transferable between similar molecules.[26,27] Assuming that this transferability of the energies of the partitioned atoms applies to the species HA and A^-,

$$E_{A^-} - E_{HA} \approx -E_H$$

where E_H is the AIM energy of the ionizable proton in the acid HA, the AIM energies of all the other atoms canceling. Thus

$$pK_a \approx \frac{-E_H}{2.303RT} + C \quad (5)$$

To obtain pK_a values in solution, the E_H values used in eq 5 are computed using wave functions obtained by the polarized continuum method, that is,

$$pK_a^{soln} \approx \frac{-E_H^{soln}}{2.303RT} + C^{soln} \quad (6)$$

This expression is the structure−property relation to be tested. This is a linear relation between pK_a^{soln} and E_H^{soln} with a slope of $-1.0/(2.303RT)$, which has a value of −0.1752 mol/kJ at 25 °C.

3. Computational Methods

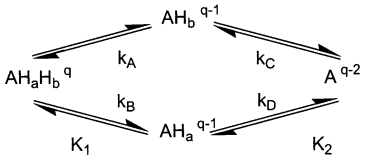
Ab initio geometry optimizations in water solution were performed with the Gaussian 98 software[28] package using the COSMO variant of the PCM solvation model,[29,30] the PW91 density functional,[31] and the all-electron 6-311+G(d,p) basis.[32] The computed E_H^{soln} values depend on the basis set used and were still not completely converged even at the largest basis investigated, 6-311+G(3df,3pd); however, provided the same basis level is used for all members of a series of acids, differences in E_H^{soln} values are converged satisfactorily. The smaller 6-311+G(d,p) basis that was finally used was chosen as a compromise between accuracy and computational expense. Likewise, different choices for the exchange−correlation functional used in the density functional procedure also give slightly different values for E_H^{soln}, and hence the same functional must always be used when comparing the E_H^{soln} values for a series of acids. The solute−solvent boundary surface in the PCM method is defined as a set of interlocking spheres with the surface of each sphere tessellated into a set of triangles. The number of triangles used for the tessellation of the spheres was increased from the default value of 60 used by Gaussian 98 to 196 where possible, giving a surface area for each triangle of <0.2 Å^2. No symmetry constraints were used for the geometry optimizations, and the coordinates of the starting structures were taken from X-ray data where available or constructed by model building using the Spartan[33] software package. Although in many cases several low-energy conformations were examined, no attempt was made to perform an exhaustive search of the potential surface to locate the global minimum. The optimized coordinates of all species are available in the Supporting Information.

The atomic basins in Bader's AIM theory are separated by surfaces defined by the condition

$$\nabla\rho(x)^T \cdot n(x) = 0$$

for all points x in the surface where $\rho(x)$ is the molecular electronic charge density, $\nabla\rho(x)$ is the gradient vector of $\rho(x)$, and $n(x)$ is the vector normal to the surface. Atomic properties

SCHEME 1: Dissociation Equilibria for Diprotic Acids



are then calculated by 3D volume integrations over the atomic basis. The kinetic energy of an atomic basin Ω is evaluated as the integral

$$K_\Omega = -\frac{1}{2}\int_\Omega \sum_i n_i \Psi_i \nabla^2 \Psi_i \, d\tau$$

where n_i is the occupation number of molecular orbital i with wave function Ψ_i. The total energy E_Ω of the atomic basin is obtained from the kinetic energy by using the virial theorem

$$E_\Omega = K_\Omega(1 - V_R)$$

where the virial ratio V_R is defined as $-V/T$ where V and T are the total potential and kinetic energies, respectively, for the entire system. The integrations over the atomic basins were performed by the locally written program Dfaim that reads a wfn-type file produced by Gaussian 98. The algorithm used for the integrations in Dfaim is similar to the PROMEGA algorithm used in the program Proaimv, which is part of the Aimpac package from the Bader group.[34] This algorithm is computationally very expensive, although very robust, but is very suitable for implementation as a parallel code. The accuracy of the Dfaim program was verified by comparison with results produced by the Proaimv and Morphy98[35] programs. The parameters used for the integrations in Dfaim were such that the integration error in the energy of the ionizable hydrogen atom was always <0.1 kJ mol^{-1}. All calculations were performed on an SGI Origin 3400 system.

The acids used in this study were chosen to satisfy three criteria. First, only high-quality experimental pK_a values taken from the NIST database[36] were used, and these are at 25 °C and extrapolated to zero ionic strength unless otherwise noted. Several additional values not found in the NIST database were taken from the compilation of Serjeant and Dempsey.[37] A second criterion used to select the acids was based upon the experimental ΔH and ΔS values reported in the NIST database. An examination of these values at 25 °C shows that, with a few exceptions, the $T\Delta S$ term dominates in the standard expression for ΔG for carboxylic acids, and for anilinium and pyridinium ions, the ΔH term dominates, whereas both terms are comparable in magnitude for the phenols. Acids from each of these groups were therefore studied to explore the limits of the method. Finally, within each group, examples were chosen to span a wide range of pK_a values.

Complications arise when acids contain more than one ionizable proton, the possible equilibria for the case of a diprotic acid containing two ionizable protons being shown in Scheme 1.[38]

These equilibria involve four microscopic equilibrium constants (k_A, k_B, k_C, and k_D) and two overall macroscopic equilibrium constants (K_1 and K_2). The microscopic constants are the values obtained by computational methods based upon the molecular structure, whereas the macroscopic equilibrium constants are the values obtained from experiment. Relationships

between the macroscopic and microscopic constants are easily obtained.[38]

$$K_1 = k_A + k_B \tag{7}$$

$$\frac{1}{K_2} = \frac{1}{k_C} + \frac{1}{k_D} \tag{8}$$

$$\frac{k_A}{k_B} = \frac{k_D}{k_C} \tag{9}$$

If the microscopic values pk_A and pk_B and also pk_C and pk_D differ by more than about 2 pK_a units,

$$pK_1 \approx \min(pk_A, pk_B)$$

$$pK_2 \approx \max(pk_C, pk_D)$$

otherwise eqs 7, 8, and 9 must be used to relate the two sets of constants, and in this case, computations of the AIM energies of the ionizable protons for the species $AK_aH_b^q$ and either AH_a^{q-1} or AH_b^{q-1} are then required. An important special case arises when the two ionizable protons are identical in, for example, dicarboxylic acids, where $k_A = k_B$ and $k_C = k_D$. Equations 7 and 8 then give pK_1 = pk_A − log 2 and pK_2 = pk_C + log 2. Although the size of this effect is at the limit of the accuracy of the method, these corrections have been applied to the experimental macroscopic values used in this study before comparison with the computed microscopic values.

The proposed structure−property relation, eq 6, was tested by performing standard least-squares linear regressions of experimentally measured pK_a^{soln} values against computed values of E_H^{soln} for various series of acids. The quality of the fit of the data to the proposed relation is judged by the R^2 value of the regression together with the average value of the unsigned errors.

4. Results and Discussion

4.1. Aliphatic Carboxylic Acids. The results for a linear regression of pK_a^{soln} against E_H^{soln} for a series of aliphatic carboxylic acids is presented in Table 1.

The R^2 value of 0.983 for the regression indicates that the proposed structure−property relation, eq 6, is obeyed very closely with an average error of only about 0.1 pK_a units. Hence, this relation can be used as a good predictor of relative pK_a values for carboxylic acids. However, the slope of the regression, −0.0873 mol/kJ, deviates considerably from the theoretical value of −0.1752 mol/kJ, and it is fruitful to investigate possible causes of this discrepancy. One reason for this difference could be that the actual structures of the acids in solution differ from those obtained from a naïve application of the PCM solvation model. This solvation model takes the effect of the solvent into account only in an average manner and does not include specific solute−solvent interactions. It might be expected that such interactions, particularly those involving the ionizable proton, could cause changes in the calculated values of E_H^{soln}. There is theoretical and experimental evidence that, at least in the gas phase, carboxylic acids can form hydrogen bonds to water molecules through the hydrogen atom of the carboxyl group. Gas-phase computational studies of the monohydrates of formic,[39−44] acetic,[41,45] and benzoic[45] acids and also of the dihydrate and higher hydrates of formic acid[42−44,46] have found stable hydrogen-bonded complexes involving the ionizable proton in the bonding. The mono- and dihydrates of formic acid have also been investigated experimentally by means of microwave spectroscopy.[43] The consensus of these investigations, both experimental and computational, is that in the gas phase the most stable form of the monohydrate has a planar six-membered ring involving two hydrogen bonds whereas the dihydrate adopts a planar eight-membered ring structure involving three hydrogen bonds. These general structures for the formic acid hydrates are shown in Chart 1.

This indicates that structures involving water molecules hydrogen bonded to the carboxyl group may be particularly stable. In this work, the effect of including water molecules hydrogen bonded to the carboxyl group has been investigated by forming supermolecules involving the acid solute and either one or two water molecules using the gas-phase structures described in Chart 1. These structures were then used as the starting point for full solution-phase PCM geometry optimizations of the supermolecules. No attempt was made specifically to include water molecules that may be interacting with other parts of the solute, and indeed Cramer and Truhlar in their review[14] have discussed some of the difficulties that may be encountered when attempts are made to include explicitly more weakly bound solvent molecules from the first solvation shell. In all cases investigated, stable hydrogen-bonded hydrate structures were found in solution; however, the ring structure of the monohydrates is disrupted with only the hydrogen bond involving the ionizable proton remaining. In contrast, the ring

TABLE 1: Aliphatic Carboxylic Acids[a]

acid	pK_a[b] expt	E_H kJ/mol	pK_a[c] calc	error
malonic pK_{a2}	5.395[d]	−910.18	5.379	−0.016
succinic pK_{a2}	5.335[d]	−908.76	5.255	−0.080
trimethylacetic	5.032	−904.06	4.844	−0.188
propanoic	4.874	−904.96	4.923	0.049
butanoic	4.818	−905.67	4.985	0.167
acetic	4.757	−902.94	4.747	−0.010
succinic pK_{a1}	4.508[e]	−899.79	4.472	−0.036
acrylic	4.258	−898.50	4.359	0.101
fumaric pK_{a2}	4.18[d]	−898.71	4.377	0.197
3-chloropropanoic	4.11	−894.56	4.015	−0.095
glycolic	3.832	−892.28	3.816	−0.016
formic	3.744	−892.04	3.795	0.051
fumaric pK_{a1}	3.32[e]	−885.53	3.227	−0.093
bromoacetic	2.902	−883.40	3.041	0.139
2-chloropropanoic	2.90	−885.71	3.242	0.342
chloroacetic	2.862	−883.11	3.016	0.154
fluoroacetic	2.586	−878.28	2.594	0.008
cyanoacetic	2.472	−877.07	2.488	0.016
maleic pK_{a1}	1.92	−868.23	1.716	−0.204
regression parameters:				
R^2			0.983	
m, mol/kJ			−0.0873	
c			74.080	
mean unsigned error			0.103	
max unsigned error			0.342	

[a] Regression: pK_a = $m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation. [d] Macroscopic experimental value converted to microscopic value by subtraction of log 2; see text. [e] Macroscopic experimental value converted to microscopic value by addition of log 2; see text.

CHART 1: Mono- and Dihydrates of Formic Acid

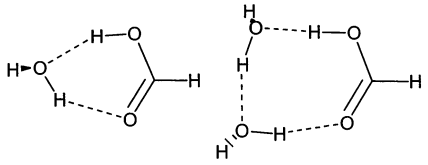


TABLE 2: Aliphatic Carboxylic Acid Monohydrates[a]

acid	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
malonic pK_{a2}	5.395[d]	−913.78	5.363	−0.032
succinic pK_{a2}	5.335[d]	−911.63	5.050	−0.285
trimethylacetic	5.032	−910.02	4.816	−0.216
propanoic	4.874	−910.52	4.889	0.015
butanoic	4.818	−910.68	4.912	0.094
acetic	4.757	−909.63	4.759	0.002
succinic pK_{a1}	4.508[e]	−908.45	4.587	0.079
acrylic	4.258	−906.90	4.362	0.104
fumaric pK_{a2}	4.18[d]	−904.59	4.025	−0.155
3-chloropropanoic	4.11	−904.41	3.999	−0.111
glycolic	3.832	−900.60	3.444	−0.388
formic	3.744	−904.04	3.945	0.201
fumaric pK_{a1}	3.32[e]	−898.37	3.120	−0.200
bromoacetic	2.902	−895.90	2.760	−0.142
2-chloropropanoic	2.90	−898.97	3.207	0.307
chloroacetic	2.862	−895.90	2.760	−0.102
fluoroacetic	2.586	−893.43	2.400	−0.186
cyanoacetic	2.472	−894.64	2.577	0.105
maleic pK_{a1}	1.92	−890.54	1.980	0.060
regression parameters:				
R^2			0.972	
m, mol/kJ			−0.1456	
c			−127.683	
mean unsigned error			0.146	
max unsigned error			0.388	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation. [d] Macroscopic experimental value converted to microscopic value by subtraction of log 2; see text. [e] Macroscopic experimental value converted to microscopic value by addition of log 2; see text.

TABLE 3: Aliphatic Carboxylic Acid Dihydrates[a]

acid	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
malonic pK_{a2}	5.395[d]	−905.19	5.548	0.153
succinic pK_{a2}	5.335[d]	−902.73	5.112	−0.223
trimethylacetic	5.032	−900.81	4.772	−0.260
propanoic	4.874	−901.33	4.864	−0.010
butanoic	4.818	−901.68	4.926	0.108
acetic	4.757	−900.63	4.741	−0.016
succinic pK_{a1}	4.508[e]	−898.87	4.429	−0.079
acrylic	4.258	−898.45	4.355	0.097
fumaric pK_{a2}	4.18[d]	−897.29	4.149	−0.031
3-chloropropanoic	4.11	−895.82	3.889	−0.221
glycolic	3.832	−894.93	3.732	−0.100
formic	3.744	−896.92	4.084	0.340
fumaric pK_{a1}	3.32[e]	−890.78	2.997	−0.323
bromoacetic	2.902	−890.31	2.914	0.012
2-chloropropanoic	2.90	−892.78	3.351	0.451
chloroacetic	2.862	−891.23	3.077	0.215
fluoroacetic	2.586	−888.92	2.668	0.082
cyanoacetic	2.472	−889.31	2.737	0.265
maleic pK_{a1}	1.92	−883.11	1.639	−0.281
regression parameters:				
R^2			0.957	
m, mol/kJ			−0.1770	
c			−154.671	
mean unsigned error			0.172	
max unsigned error			0.451	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation. [d] Macroscopic experimental value converted to microscopic value by subtraction of log 2; see text. [e] Macroscopic experimental value converted to microscopic value by addition of log 2; see text.

structure involving the two hydrogen-bonded water molecules is retained in the dihydrates. The solution-phase optimized structures of the hydrated supermolecules are included in the Supporting Information. Regression results for the monohydrate and dihydrate acids are presented in Tables 2 and 3, respectively, for the same set of compounds used previously.

Again, it can be seen that the proposed structure−property relation is obeyed very well, as indicated by the R^2 values and average errors, with the slope of the regression for the dihydrate, −0.1770 mol/kJ, approaching the theoretical value of −0.1752 mol/kJ very closely, indicating that the dihydrate is probably the predominant species involved in the proton dissociation process. The geometry optimization of some of the hydrated acids proved to be more difficult than for the corresponding unhydrated species, probably because of a flatter potential surface associated with the hydrogen-bonded structures. This is a possible reason for slightly more scatter in the computed pK_a values for the hydrated acids compared to those for the unhydrated acids.

Calculated pK_a values for the dihydrates of several acids that were not included in the previous tables are given in Table 4. These values, calculated using the structure−property relation together with the regression parameters of Table 3, require some further comment.

In contrast to the other diprotic acids investigated, the two ionizable protons in maleic acid are not equivalent since one of them is involved in an internal hydrogen bond, and as a result, the correction for statistical factors does not have to be applied to the experimental pK_a values. The computations indicate that, as might be expected, the proton involved in the first dissociation is the one not involved in the internal hydrogen bonding, and the calculated pK_{a1} values for maleic acid shown in Tables 1, 2, and 3 for structures with zero, one, and two waters of hydration, respectively, fit the respective regression relations

TABLE 4: Aliphatic Carboxylic Acid Dihydrates

acid	pK_a[a] exptl	E_H kJ/mol	pK_a[b] calcd	error
maleic pK_{a2}	6.27	−908.16	6.073	−0.197
oxalic pK_{a2}	3.965[c]	−900.57	4.730	0.765
oxalic pK_{a1}	1.551[d]	−886.82	2.296	0.745
phenylacetic	4.31	−901.23	4.847	0.537
malonic pK_{a1}	3.148[c]	−895.74	3.875	0.727
glyoxylic	3.46			
keto form		−885.63	2.086	−1.374
diol form		−893.48	3.475	0.015
pyruvic	2.48			
keto form		−888.65	2.620	0.14
diol form		−891.59	3.140	0.66

[a] Values taken from ref 36 (25 °C, zero ionic strength). [b] Computed from the regression relation of Table 3. [c] Macroscopic experimental value converted to microscopic value by subtraction of log 2; see text. [d] Macroscopic experimental value converted to microscopic value by addition of log 2; see text.

well. When calculating pK_{a2} for maleic acid, it was found that the internal hydrogen bond was maintained in the lowest-energy structure of the unhydrated hydrogen maleate ion. In contrast, the internal hydrogen bond is replaced by external hydrogen bonds to water molecules in the lowest-energy structures of the hydrated forms. The calculated pK_{a2} value for the externally hydrogen-bonded dihydrated species is shown in Table 4 and agrees reasonably well with the experimental value. A similar effect arises with the pK_{a1} and pK_{a2} values for oxalic acid. Here, both ionizable protons are equivalent and are involved in internal hydrogen bonds in the most stable structure for unhydrated oxalic acid, whereas in the hydrated structures, again it was found that the internal hydrogen bonds are replaced by external hydrogen bonds to water molecules. However, in this case, both of the calculated pK_{a1} and pK_{a2} values deviate somewhat from

TABLE 5: 3,4-Substituted Benzoic Acid Dihydrates[a]

substituent	pK_a[a] exptl	E_H kJ/mol	pK_a[b] calcd	error
4-OCH_3	4.475	−898.21	4.520	0.045
4-CH_3	4.370	−896.21	4.237	−0.133
3-CH_3	4.273	−896.03	4.212	−0.061
H	4.202	−896.08	4.219	0.017
3-COO^-	4.20[c]	−895.11	4.082	−0.118
4-COO^-	4.18[c]	−895.01	4.067	−0.113
4-F	4.131	−894.06	3.933	−0.198
3-OCH_3	4.090	−896.08	4.219	0.129
4-Cl	3.986	−894.22	3.956	−0.030
4-COOH·$2H_2O$	3.88[d]	−892.62	3.729	−0.151
3-F	3.863	−892.17	3.666	−0.197
3-Cl	3.825	−892.72	3.744	−0.081
3-COOH·$2H_2O$	3.80[d]	−893.62	3.871	0.071
3-CN	3.596	−892.14	3.662	0.066
4-CN	3.550	−890.88	3.483	−0.067
3-NO_2	3.449	−891.36	3.551	0.102
4-NO_2	3.442	−891.28	3.540	0.098
regression parameters:				
R^2			0.886	
m, mol/kJ			−0.1414	
c			−122.487	
mean unsigned error			0.099	
max unsigned error			0.198	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation. [d] Macroscopic experimental value converted to microscopic value by subtraction of log 2; see text. [e] Macroscopic experimental value converted to microscopic value by addition of log 2; see text.

the experimental values. Phenylacetic acid is another outlier, as is the pK_{a1} value for malonic acid, and for this reason, these values were not included in the regression analyses. However, the pK_{a2} value of malonic acid is satisfactory.

Calculations for the α keto acids—glyoxylic and pyruvic—illustrate another point. It is possible for both of these species to add a water molecule to the carbonyl group to form a diol, and hence both the keto and diol structures have been investigated. The results in Table 4 show clearly that in the case of glyoxylic acid the diol structure accounts for the observed pK_a value, whereas the keto structure is the preferred one for pyruvic acid. These conclusions are of course not new but serve to illustrate the potential of this technique for the investigation of the structures of acid species in solution.

4.2. Substituted Benzoic Acids. Calculations have been carried out for various substituted benzoic acids with zero, one, or two water molecules hydrogen bonded to the carboxyl group in the same manner as for the aliphatic carboxylic acids. Again, all three types of hydrated species obey the proposed structure−property relationship closely, but only the slope of the regression for the dihydrates approaches the theoretical value. The results for the dihydrates are presented in Table 5, whereas the results for the unhydrated and monohydrated benzoic acids are provided in the Supporting Information. Table 6 shows the results for the dihydrates of some additional benzoic acids not included in the regression analysis.

The 4-CHO- and 3-CHO-substituted benzoic acids were not included in the regression analysis because the ionic strength at which the experimental pK_a values were obtained is uncertain; however, despite this, the calculated pK_a values appear to be very satisfactory. The values calculated for the 2-substituted benzoic acids have larger errors appearing to be systematically overestimated, and Table 7 shows some of these values fitted to a regression equation independently of those for the 3- and 4-substituted acids.

TABLE 6: Substituted Benzoic Acid Dihydrates

substituent	pK_a[a] exptl	E_H kJ/mol	pK_a[b] calcd	error
3-CHO	3.84[c]	−892.75	3.748	−0.092
4-CHO	3.75[c]	−891.44	3.563	−0.187
2-OCH_3	4.094	−900.00	4.773	0.679
2-CH_3	3.903	−897.32	4.394	0.491
2-F	3.469	−893.25	3.819	0.350
2-Cl	2.925	−892.64	3.732	0.807
2-NO_2	2.185	−891.15	3.522	1.337

[a] Values taken from ref 36 (25 °C, zero ionic strength) unless otherwise noted. [b] Computed from the regression relation of Table 5. [c] Values taken from ref 37 (25 °C, ionic strength uncertain).

TABLE 7: 2-Substituted Benzoic Acid Dihydrates[a]

substituent	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
2-OCH_3	4.094	−900.00	4.188	0.094
2-CH_3	3.903	−897.32	3.820	−0.083
2-F	3.469	−893.25	3.261	−0.208
2-Cl	2.925	−892.64	3.177	0.252
regression parameters:				
R^2			0.849	
m, mol/kJ			−0.1373	
c			−119.382	
mean unsigned error			0.159	
max unsigned error			0.252	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation.

Although only four data points were used in the regression for the 2-substituted benzoic acids, giving a somewhat poorer fit than for previous regressions, it seems clear that the regression parameters for the 2-substituted benzoic acids are slightly different from those for the 3- and 4-substituted acids. The pK_a of 2-nitrobenzoic acid is not predicted satisfactorily by either one of these relations, with the calculated value having an error of 0.788 using the 2-substituted regression and an error of 1.337 from the 3,4-substituted regression. In this compound, both the carboxyl and nitro groups are rotated out of the plane of the aromatic ring presumably because of large steric interactions between the two bulky substituent groups, whereas in the other substituted benzoic acids, the carboxyl group is coplanar with the aromatic ring. It is also interesting that it has long been known that the 2-substituted benzoic acids do not obey the Hammett relations obeyed by the 3- and 4-substituted species.[47]

4.3. Substituted Phenols. In the case of phenol, there is again theoretical and spectroscopic evidence for the formation of hydrogen-bonded complexes between phenol and various-sized water clusters in the gas phase. Complexes of phenol with clusters of one,[48] two,[49,50] three,[51] and more[50,52] water molecules have been studied, and these investigations indicate that one water molecule is closely bound to the phenol group with the other water molecules at a greater distance. These gas-phase geometries were used as starting points for preliminary solution-phase investigations of phenols with zero, one, and two bound water molecules. These calculations indicated that while increasing the number of hydrogen-bonded water molecules from zero to one produced a significant change in the regression slope, an increase from one to two bound waters produced little further change with many of the dihydrates in fact proving to be unstable at the level of theory used, with the more weakly bound water leaving the complex. Only detailed results for some monohydrated phenols are reported in Table 8 , and the details

TABLE 8: Substituted Phenol Monohydrates[a]

substituent	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
2-CH_3	10.31	−931.74	10.188	−0.122
4-CH_3	10.269	−932.71	10.337	0.068
4-OCH_3	10.21	−933.92	10.524	0.314
3-CH_3	10.095	−931.24	10.111	0.016
H	9.997	−930.79	10.042	0.045
2-OCH_3	9.98	−927.62	9.553	−0.427
4-F	9.908	−929.16	9.791	−0.117
3-OCH_3	9.652	−929.43	9.832	0.180
4-Cl	9.426	−927.51	9.537	0.111
3-F	9.206	−924.33	9.047	−0.159
3-Cl	9.125	−923.83	8.970	−0.155
3-CHO	8.988	−923.65	8.942	−0.046
2-F	8.705	−918.22	8.106	−0.599
3-CN	8.57	−921.13	8.554	−0.016
2-Cl	8.531	−919.16	8.251	−0.280
2-CHO	8.374	−921.47	8.606	0.232
3-NO_2	8.37	−918.56	8.158	−0.212
4-CN	7.97	−919.69	8.332	0.362
2-NO_2	7.230	−915.51	7.689	0.459
4-NO_2	7.149	−913.52	7.382	0.233
regression parameters:				
R^2			0.926	
m, mol/kJ			−0.1540	
c			−133.300	
mean unsigned error			0.208	
max unsigned error			0.599	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation.

TABLE 9: Substituted Anilinium Ions (Unhydrated)[a]

substituent	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
4-OCH_3	5.357	−1102.87	5.302	−0.055
4-CH_3	5.080	−1100.06	4.955	−0.125
3-CH_3	4.712	−1099.19	4.848	0.136
4-F	4.654	−1096.46	4.511	−0.143
H	4.601	−1098.33	4.742	0.141
2-OCH_3	4.526	−1098.06	4.708	0.182
3-OCH_3	4.204	−1096.17	4.475	0.271
4-Cl	3.979	−1093.00	4.084	0.105
3-F	3.588	−1088.93	3.581	−0.007
3-Cl	3.520	−1087.64	3.422	−0.098
2-F	3.203	−1084.52	3.036	−0.167
3-NO_2	2.460	−1082.02	2.727	0.267
regression parameters:				
R^2			0.963	
m, mol/kJ			−0.1235	
c			−130.902	
mean unsigned error			0.141	
max unsigned error			0.271	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation.

for the corresponding unhydrated species are included with the Supporting Information.

The R^2 values show that the structure−property relation is fitted well with the regression slope for the monohydrates again approaching the theoretical value. Eliminating the 2-substituted phenols to fit them to a separate regression relation produced only a marginal improvement, and so they are retained.

4.4. Substituted Anilinium Ions. A regression analysis for some substituted unhydrated anilinium ions is presented in Table 9, and the fit to the regression relation is reasonable.

In general, the geometry optimizations for these species proved difficult, and in most cases, several local minima of very similar energies were found. Without a more systematic search of the potential surface, it is not possible to be certain that the global minimum was always found or whether in fact a conformational average should be used. Another complication is the ambiguity as to which of the hydrogen atoms attached to the nitrogen atom is to be selected for the evaluation of E_H^{soln}. In this work, the proton with the lowest electron density at the bond critical point in the electron density between the nitrogen and hydrogen atoms was chosen on the basis of the assumption that this corresponds to the weakest bond.

TABLE 10: Substituted Anilinium Ions (Unhydrated)

substituent	pK_a[a] exptl	E_H kJ/mol	pK_a[b] calcd	error
4-NH_2	5.76[c,d]	−1110.74	6.274	0.514
3-NH_2	4.81[c,d]	−1101.40	5.121	0.311
4-NH_3^+	3.17[d,e]	−1082.55	2.793	−0.377
3-NH_3^+	2.80[d,e]	−1080.68	2.562	−0.238
2-Cl	2.653	−1086.88	3.328	0.675
2-CH_3	4.447	−1099.80	4.923	0.476
2-NH_2	4.36[c,d]	−1076.43	2.037	−2.323
2-NH_3^+	0.9[e,f]	−1077.90	2.219	1.319
2-NO_2	−0.25	−1082.94	2.841	3.091
4-NO_2	1.015	−1081.36	2.646	1.631

[a] Values taken from ref 36 (25 °C, zero ionic strength) unless otherwise noted. [b] Computed from the regression relation of Table 9. [c] Macroscopic experimental value converted to microscopic value by subtraction of log 2; see text. [d] Ionic strength 0.1. [e] Macroscopic experimental value converted to microscopic value by addition of log 2; see text. [f] Ionic strength 0.5.

Some other data not included in the regression analysis is shown in Table 10, with the calculated values obtained using the regression parameters of Table 9. This Table includes data for the species with substituents, R = 3NH_2, 4NH_2, 3NH_3^+ and 4NH_3^+, which have experimental values reported at nonzero ionic strengths, but nevertheless the agreement between these experimental values and the calculated values is reasonable. The results for the 2-substituted anilinium ions are somewhat ambiguous. The 2-OCH_3- and 2-F-substituted ions obey the regression equation well, whereas the 2-Cl- and 2-CH_3-substituted species show larger deviations. The calculated pK_a values for the 2-NH_2- and 2-NO_2-substituted ions are also grossly in error; however, it was judged that an insufficient number of accurate results are available for an investigation of the possibility of a separate regression equation for the 2-substituted ions.

Some preliminary studies of the hydration behavior of anilinium ions were made using complexes having one water molecule hydrogen bonded to each proton of the anilinium group. Although these complexes were found to be stable, the optimization of the geometries proved to be difficult and time-consuming. In addition, the regression relation appeared to have a slope very similar to that for the unhydrated ions, and so a more detailed investigation of the hydrated species was not pursued.

4.5. Substituted Pyridinium Ions. It has been shown both experimentally and computationally that in the gas phase pyridine forms a hydrogen-bonded monohydrate with the water acting as the proton donor.[53−55] Accordingly, in this work, similar monohydrates of substituted pyridinium ions, with the pyridiniun acting as the proton donor, were examined in aqueous solutions. The monohydrates were found to be stable, and eq 6 has been tested for both unhydrated and monohydrated substituted pyridinium ions. The regression for the 3- and 4-substituted monohydrates is shown in Table 11, whereas the data for the unhydrated species is provided in the Supporting Information.

TABLE 11: 3,4-Substituted Pyridinium Monohydrates[a]

substituent	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
4-OCH_3	6.470	−1049.62	6.287	−0.183
4-CH_3	6.00	−1044.58	5.561	−0.439
3-CH_3	5.68	−1048.47	6.122	0.442
H	5.20	−1040.75	5.009	−0.191
4-Br	3.68	−1029.06	3.325	−0.355
3-Br	2.87	−1026.65	2.977	0.107
3-Cl	2.84	−1027.99	3.170	0.330
4-CN	1.48	−1017.20	1.616	0.136
4-NO_2	1.23	−1011.50	0.794	−0.436
3-CN	1.17	−1018.85	1.853	0.683
3-NO_2	0.80	−1009.53	0.510	−0.290
regression parameters:				
R^2			0.968	
m, mol/kJ			−0.1441	
c			−144.963	
mean unsigned error			0.327	
max unsigned error			0.683	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation.

TABLE 12: 2-Substituted Pyridinium Monohydrates[a]

substituent	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
2-CH_3	5.95	−1057.81	5.946	−0.004
2-Br	0.70	−1018.48	0.730	0.030
2-Cl	0.49	−1016.41	0.456	−0.034
regression parameters:				
R^2			0.999	
m, mol/kJ			−0.1326	
c			−134.320	
mean unsigned error			0.023	
max unsigned error			0.034	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation.

The unhydrated species were found to give a poor fit to the proposed structure−property relation (R^2 value of 0.780), whereas the monohydrates fit the proposed relation much better, with the regression slope of −0.1441 mol/kJ approaching the theoretical value. Here again it appears necessary to fit the 2-substituted species to a different regression, as shown in Table 12, although with only three values in the set it is premature to draw firm conclusions.

4.6. Combined Regression Analysis of Aliphatic Carboxylic Acids, Substituted Benzoic Acids, and Substituted Phenols. The regression parameters found for the dihydrated carboxylic and benzoic acids and monohydrated phenols in Tables 3, 5, and 8 are similar, and so the data from these Tables was used in a combined regression analysis, as shown in Table 13.

As indicated by the average error, there is slightly more scatter of the data points in the combined fit compared to that in the fits of the individual data sets; nevertheless, the fit is good, and the regression slope for the combined data is very close to the theoretical value. By contrast, the combined fit for the unhydrated species, included in the Supporting Information, is poorer. It was also found that the data for the substituted anilinium and pyridinium ions could not be included in the regression of Table 13. This is not surprising in view of the approximations made in the "derivation" of eq 6 since all of the data in Table 13 refers to the breaking of O−H bonds whereas N−H bonds are broken in the dissociation of the anilinium and pyridinium ions.

TABLE 13: Aliphatic Carboxylic Acid Dihydrates, 3,4-Substituted Benzoic Acid Dihydrates, and Substituted Phenol Monohydrates[a]

acid	pK_a[b] exptl	E_H kJ/mol	pK_a[c] calcd	error
2-CH_3 phenol	10.31	−931.74	10.239	−0.071
4-CH_3 phenol	10.269	−932.71	10.406	0.137
4-OCH_3 phenol	10.21	−933.92	10.615	0.405
3-CH_3 phenol	10.095	−931.24	10.153	0.058
phenol	9.997	−930.79	10.075	0.078
2-OCH_3 phenol	9.98	−927.62	9.528	−0.452
4-F phenol	9.908	−929.16	9.794	−0.114
3-OCH3 phenol	9.652	−929.43	9.841	0.189
4-Cl phenol	9.426	−927.51	9.509	0.083
3-F phenol	9.206	−924.33	8.961	−0.245
3-Cl phenol	9.125	−923.83	8.875	−0.250
3-CHO phenol	8.988	−923.65	8.844	−0.144
3-CN phenol	8.57	−921.13	8.409	−0.161
2-Cl phenol	8.531	−919.16	8.069	−0.462
2-CHO phenol	8.374	−921.47	8.468	0.094
3-NO_2 phenol	8.37	−918.56	7.966	−0.404
4-NO_2 phenol	7.149	−913.52	7.096	−0.053
malonic pK_{a2}	5.395[d]	−905.19	5.659	0.264
succinic pK_{a2}	5.335[d]	−902.73	5.235	−0.100
trimethylacetic	5.032	−900.81	4.904	−0.128
propanoic	4.874	−901.33	4.993	0.119
butanoic	4.818	−901.68	5.054	0.236
acetic	4.757	−900.63	4.873	0.116
succinic pK_{a1}	4.508[e]	−898.87	4.569	0.061
4-OCH_3 benzoic	4.475	−898.21	4.455	−0.020
4-CH_3 benzoic	4.370	−896.21	4.110	−0.260
3-CH_3 benzoic	4.273	−896.03	4.079	−0.194
acrylic	4.258	−898.45	4.497	0.239
benzoic	4.202	−896.08	4.088	−0.114
3-COO− benzoic	4.20[d]	−895.11	3.920	−0.280
fumaric pK_{a2}	4.18[d]	−897.29	4.297	0.117
4-COO− benzoic	4.18[d]	−895.01	3.903	−0.277
4-F benzoic	4.131	−894.06	3.739	−0.392
3-chloropropanoic	4.11	−895.82	4.043	−0.067
3-OCH_3 benzoic	4.090	−896.08	4.088	−0.002
4-Cl benzoic	3.986	−894.22	3.767	−0.219
4-COOH benzoic	3.88[e]	−892.62	3.491	−0.389
3-F benzoic	3.863	−892.17	3.413	−0.450
glycolic	3.832	−894.93	3.889	0.057
3-Cl benzoic	3.825	−892.72	3.508	−0.317
3-COOH benzoic	3.80[e]	−893.62	3.663	−0.137
formic	3.744	−896.92	4.233	0.489
3-CN benzoic	3.596	−892.14	3.408	−0.188
4-CN benzoic	3.55	−890.88	3.191	−0.359
3-NO_2 benzoic	3.449	−891.36	3.274	−0.175
4-NO_2 benzoic	3.442	−891.28	3.260	−0.182
fumaric pK_{a1}	3.32[e]	−890.78	3.174	−0.146
bromoacetic	2.902	−890.31	3.092	0.190
2-chloropropanoic	2.90	−892.78	3.519	0.619
chloroacetic	2.862	−891.23	3.251	0.389
fluoroacetic	2.586	−888.92	2.853	0.267
cyanoacetic	2.472	−889.31	2.920	0.448
maleic pK_{a1}	1.92	−883.11	1.850	−0.070
regression parameters:				
R^2			0.991	
m, mol/kJ			−0.1725	
c			−150.486	
mean unsigned error			0.217	
max unsigned error			0.619	

[a] Regression p$K_a = m \times E_H + c$. [b] Values taken from ref 36 (25 °C, zero ionic strength). [c] Computed from the regression relation. [d] Macroscopic experimental value converted to microscopic value by subtraction of log 2; see text. [e] Macroscopic experimental value converted to microscopic value by addition of log 2; see text.

4.7. Computation of pK_a Values for Some Diprotic Acids. The possible dissociation pathways in diprotic acids are shown in Scheme 1, with the relations between the macroscopic and microscopic dissociation constants given by eqs 7 and 8. Some

TABLE 14: Computed Microscopic pK_a Values for Some Benzenes with Two Substituent Groups R_a and R_b

parent acid	R_a	R_b	E_{Ha} kJ/mol	E_{Hb} kJ/mol	pk_A	pk_B	pk_C	pk_D
4-hydroxy benzoic	COOH	4OH	−898.26	−922.50	4.46[a]	8.65[b]		
	COO^-	4OH		−927.75			9.55[b]	
	COOH	$4O^-$	−908.37					6.21[a]
3-hydroxy benzoic	COOH	3OH	−895.95	−927.27	4.07[a]	9.47[b]		
	COO^-	3OH		−930.82			10.08[b]	
	COOH	$3O^-$	−901.57					5.04[a]
4-amino benzoic	COOH	$4NH_3^+$	−891.33	−1090.24	3.27[a]	3.74[c]		
	COO^-	$4NH_3^+$		−1095.67			4.41[c]	
	COOH	$4NH_2$	−901.10					4.95[a]
3-amino benzoic	COOH	$3NH_3^+$	−890.81	−1090.90	3.18[a]	3.82[c]		
	COO^-	$3NH_3^+$		−1098.38			4.75[c]	
	COOH	$3NH_2$	−897.00					4.25[a]
4-pyridine carboxylic	pyridinium	4COOH	−1029.96	−882.19	3.45[d]	1.69[b]		
	pyridine	4COOH		−891.54			3.31[b]	
	pyridinium	$4COO^-$	−1046.18					5.79[d]
3-pyridine carboxylic	pyridinium	3COOH	−1033.50	−885.19	3.96[d]	2.21[b]		
	pyridine	3COOH		−893.72			3.68[b]	
	pyridinium	$3COO^-$	−1050.75					6.45[d]

[a] Value computed from E_{Ha} and the regression relation of Table 13. [b] Value computed from E_{Hb} and the regression relation of Table 13. [c] Value computed from E_{Hb} and the regression relation of Table 9. [d] Value computed from E_{Ha} and the regression relation of Table 11.

TABLE 15: Computed Macroscopic pK_a Values for Some Disubstituted Benzenes

acid	pk_A − pk_B	pk_D − pk_C	pK_1[a] exptl	pK_1[b] calcd	error	pK_2[a] exptl	pK_2[c] calcd	error
4-hydroxybenzoic	−4.19	−3.34	4.581	4.46	−0.12	9.46	9.55	0.09
3-hydroxybenzoic	−5.40	−5.04	4.080	4.07	−0.01	9.90	10.08	0.18
4-aminobenzoic	−0.47	0.54	2.41	3.14	0.73	4.87	5.06	0.19
3-aminobenzoic	−0.64	−0.50	3.08	3.09	0.01	4.76	4.87	0.11
4-pyridinecarboxylic	1.76	2.48	1.72	1.68	−0.04	4.88	5.79	0.91
							(5.07)[d]	(0.19)
3-pyridinecarboxylic	1.75	2.77	2.03	2.20	0.17	4.82	6.45	1.63
							(5.43)[d]	(0.61)

[a] Values taken from ref 36 (25 °C, zero ionic strength). [b] Computed from pk_A and pk_B using eq 7. [c] Computed from pk_C and pk_D using eq 8. [d] Computed from pk_A, pk_B, and pk_C using eqs 9 and 8, respectively.

examples of diprotic acids with identical acid groups have already been presented in previous sections, and in this section, some results are presented for some diprotic acids with nonidentical acid groups. The details of the computation of the four microscopic pk_a values (pk_A, pk_B, pk_C, and pk_D) of some 3,4-disubstituted benzenes based upon the regression parameters derived in Tables 9,11, and 13 are presented in Table 14. Note that for the acids in this Table all −COOH groups are treated as dihydrates while all −OH and pyridinium groups are monohydrates.

It should be noted that because of eq 9 calculations are required only for the species $AK_aH_b^q$ and either AH_a^{q-1} or AH_b^{q-1}; however, calculations are reported here for all three species to allow for an additional consistency check. The macroscopic values pK_1 and pK_2 computed using eqs 7 and 8 are presented in Table 15 together with the values of the differences (pk_A − pk_B) and (pk_D − pk_C), which should be equal according to relation 9.

In general, the experimental macroscopic pK_a values are predicted accurately, the notable exceptions being the pK_2 values for the pyridine carboxylic acids when the directly computed microscopic pk_D values are used. In both of these cases, the values of pk_D were derived directly from computations involving a zwitterion structure, whereas values derived indirectly by the use of eq 9 together with calculations of pk_A, pk_B, and pk_C for non-zwitterion species are in much better agreement with experiment. This possibly indicates that the direct treatment of zwitterion species requires a more elaborate protocol involving larger basis sets than those used in the current work. The internal consistency check provided by the requirement that pk_A − pk_B = pk_D − pk_C is obeyed only to within about 1 pK_a unit for all of these acids.

5. Conclusions

A simple linear relation between pK_a and the energy E_H^{soln} of the dissociating proton, as calculated by Bader's atoms in molecules theory, has been obtained starting from a standard expression for the equilibrium constant. This relation has been tested by solvated density functional calculations on a number of different types of organic acids, and in all cases investigated, a good linear relation has been found between reliable experimental pK_a values and E_H^{soln}. This allows for predictions of relative pK_a values with average errors of usually less than about 0.2 pK_a units. In all cases, apart from the anilinium ions, the inclusion of specifically hydrogen-bonded water molecules greatly improved the agreement between the slope of the regression and the theoretically derived value of $-1.0/2.303RT$, indicating that these hydrogen-bonded waters play an important role in the mechanism of the dissociation process. This observation together with the calculation of pK_a values for possible alternative structures of glyoxylic and pyruvic acids indicates the possibility that this method may be developed into a useful probe of the structure of acids in solution. It should also be noted that the dissociation of the proton in the dihydrated carboxylic acids leads to a disruption of the highly ordered eight-membered ring system involving the two waters of hydration. This is consistent with the experimental observation mentioned previously that the entropy term dominates the free-energy change for the dissociation process in these systems.

However, an intriguing question still left unanswered by this work is why a method that involves only the computation of an energy quantity, E_H^{soln}, is able to describe both the entropy-driven dissociations of the carboxylic and benzoic acids and the enthalpy-driven dissociation of the anilinium and pyridinium ions.

Acknowledgment. I thank Dr. Ian Atkinson for many useful discussions and James Cook University for internal research grants as well as the provision of high-performance computer time.

Supporting Information Available: Cartesian coordinates of the optimized solution structures of all acids appearing in this study together with tables containing some additional regressions. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley & Sons: New York, 1986.

(2) Atkins, P. W. *Physical Chemistry*; Oxford University Press: Oxford, U.K., 1994

(3) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.

(4) Richardson, W. H.; Peng, C.; Bashford, D.; Noodleman, J.; Case, D. A. *Int. J. Quantum Chem.* **1997**, *61*, 207.

(5) Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.

(6) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194.

(7) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402.

(8) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J. Chem. Phys.* **2001**, *114*, 4595.

(9) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314.

(10) Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2001**, *85*, 727.

(11) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 5421.

(12) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413.

(13) Pliego, J. R.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.

(14) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(15) Dixon, S. L.; Jurs, P. C. *J. Comput. Chem.* **1993**, *12*, 1460.

(16) Klopman, G.; Fercu, D. *J. Comput. Chem.* **1994**, *15*, 1041.

(17) Ponec, R.; Amat, L.; Carbó-Dorca, R. *J. Phys. Org. Chem.* **1999**, *12*, 447.

(18) Popelier, P. L. A. *J. Phys. Chem. A* **1999**, *103*, 2883.

(19) Gross, K. C.; Seybold, P. G. *Int. J. Quantum Chem.* **2000**, *80*, 1107.

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

(21) Gross, K. C.; Seybold, P. G. *J. Org. Chem.* **2001**, *66*, 6919.

(22) Gross, K. C.; Seybold, P. G. *Int. J. Quantum Chem.* **2001**, *85*, 569.

(23) Klicic, J. J.; Friesner, R. A.; Liu, S.; Guida, W. C. *J. Phys. Chem A* **2002**, *106*, 1327.

(24) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(25) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Clarendon Press: Oxford, U.K., 1990.

(26) Bader, R. F. W.; Beddall, P. M. *J. Chem. Phys.* **1972**, *56*, 3320.

(27) Bader, R. F. W.; Messer, R. R. *Can. J. Chem.* **1974**, *52*, 2268.

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

(29) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.

(30) Barone, V.; Cossi, M. J. *J. Phys. Chem. A* **1998**, *102*, 1995.

(31) (a) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671. (b) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1993**, *48*, 4978. (c) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533. (d) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(32) (a) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639. (b) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650. (c) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.

(33) *Spartan*, version 5.1.1. Wavefunction, Inc.: Irvine, CA, 1998.

(34) AIMPAC software. http://www.chemistry.mcmaster.ca/aimpac/.

(35) Popelier, P. L. A. *Comput. Phys. Commun.* **1998**, *108*, 180.

(36) Martell, A. E.; Smith, R. M.; Motekaitis, R. J. *NIST Critically Selected Stability Constants of Metal Complexes*, version 6.0; NIST Standard Reference Database 46; National Institute of Standards and Technology: Gaithersburg, MD, 2001.

(37) *Ionisation Constants of Organic Acids in Aqueous Solution*; Serjeant, E. P.; Dempsey, B., Eds.; IUPAC Chemical Data Series, No. 23; Pergamon Press: New York, 1979.

(38) Perrin, D. D.; Dempsey, B.; Serjeant, E. P. pK_a *Prediction for Organic Acids and Bases*; Chapman and Hall: London, 1981, Chapter 2.

(39) Åstrand, P. O.; Karlström, G.; Engdahl, A.; Nelander, B. *J. Chem. Phys.* **1995**, *102*, 3534.

(40) Minyaev, R. M. *Chem. Phys. Lett.* **1996**, *262*, 194.

(41) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem. A* **1998**, *102*, 3782.

(42) Akiya, N.; Savage, P. E. *AIChE J.* **1998**, *44*, 405.

(43) Priem, D.; Ha, T. K.; Bauder, A. *J. Chem. Phys.* **2000**, *113*, 169.

(44) Aloisio, S.; Hintze, P. E.; Vaida, V. *J. Phys. Chem. A* **2002**, *106*, 363.

(45) Nagy, P. I.; Smith, D. A.; Alagona, G.; Ghio, C. *J. Phys. Chem.* **1994**, *98*, 486.

(46) Velardez, G. F.; Ferrero, J. C.; Beswick, J. A.; Daudey, J. P. *J. Phys. Chem. A* **2001**, *105*, 8769.

(47) Hammett, L. P. *Chem. Rev.* **1937**, *17*, 125.

(48) Schütz, M.; Bürgi, T.; Leutwyler, S. *J. Chem. Phys.* **1993**, *98*, 3763.

(49) Gerhards, M.; Kleinermanns, K. *J. Chem. Phys.* **1995**, *103*, 7392.

(50) Jacoby, C.; Roth, W.; Schmitt, M.; Janzen, C.; Spangenberg, D.; Kleinermanns, K. *J. Phys. Chem. A* **1998**, *102*, 4471.

(51) Bürgi, T.; Schütz, M.; Leutwyler, S. *J. Chem. Phys.* **1995**, *103*, 6350.

(52) Janzen, C.; Spangenberg, D.; Roth, W.; Kleinermanns, K. *J. Chem. Phys.* **1999**, *110*, 9698.

(53) Martoprawiro, M. A.; Bacskay, G. B. *Mol. Phys.* **1995**, *85*, 573.

(54) Dkhissi, A.; Adamowitz, L.; Maes, G. *J. Phys. Chem. A* **2000**, *104*, 2112.

(55) Schlücker, S.; Singh, R. K.; Asthana, B. P.; Popp, J.; Kiefer, W. *J. Phys. Chem. A* **2001**, *105*, 9983.