J. Phys. Chem. A **2008,** *112,* 775−782

Correlations and Predictions of Carboxylic Acid pK_a Values Using Intermolecular Structure and Properties of Hydrogen-Bonded Complexes

Li Tao,[†] Jun Han,[‡] and Fu-Ming Tao*

Department of Chemistry and Biochemistry, California State University, Fullerton, California 92834

Received: October 24, 2007

Density functional theory calculations have been preformed on a series of hydrogen-bonded complexes of substituted aliphatic and aromatic carboxylic acids with ammonia. Molecular properties, particularly those related to hydrogen bonding, have been carefully examined for their interdependence as well as dependence on the acidity of the acid. The bond length and stretching frequency of the hydroxyl group and the hydrogen-bond length and energy of the complex are shown to be highly correlated with each other and are linearly correlated with available literature pK_a values of the carboxylic acids. The linear correlations resulting from the fit to the available pK_a values can be used to predict the pK_a values of similar carboxylic acids. The pK_a values so predicted using the different molecular properties are highly consistent and in good agreement with the literature values. This study suggests that calculated molecular properties of hydrogen-bonded complexes allow effective and systematic prediction of pK_a values for a large range of organic acids using the established linear correlations. This approach is unique in its capability to determine the acidity of a particular functional group or the local acidity within a large molecular system such as a protein.

1. Introduction

The acidity of significant functional groups of an organic compound usually has an overriding impact on the molecular properties and the overall characteristics of an organic compound. The acidity is strongly associated with the physical, chemical, and biological properties as well as the applications of the compound. For example, acidity plays a major role in determining biologically and physically significant properties of chloro- and bromophenols, such as adsorption, solubility, mobility, and toxicity.[1−2] A quick, accurate method of determining the acidity of a functional group and further the acidity of a compound would have a wide range of applications, from environmental chemistry to pharmaceuticals.

Aside from experimental measurements, theoretical determination of the acidity of a compound has been a challenging and important objective of computational chemistry. Standard theoretical methods follow microscopic calculations of free energies based on the underlying physical chemistry of the acid dissociation process. Free energy calculations performed on various molecules have been proven to provide reasonably accurate estimates of established acidities of certain molecules, such as carboxylic acids.[3−5] However, these calculations are quite involved and the results are quite inaccurate, particularly for acids of complex structure. Moreover, acidity measurements provided by experiments and by free energy calculations both represent the overall acidity of an entire molecule. This may not be enough for certain applications in which the acidity of a specific functional group or in a relatively small region within a large polyprotic compound is critical for functioning, such as in certain proteins.[6] In such cases, acidity as measured by traditional methods is influenced by all acidic functional groups across the compound and may not directly reflect the acidity of a specific site. Clearly, a systematic approach is desirable to determine the acidity of a particular functional group within a large molecular system.

The acidity of a functional group represents its ability to release a proton to a basic solvent and is usually expressed as the acid dissociation constant K_a or its negative logarithm pK_a. The acidity is therefore dependent on the relative stability of the resulting anion with respect to its parent compound in a given solvent. Hence, the acidity reflects how strongly a proton is bound to the acid compound. In a usual theoretical calculation, acidity is determined by calculated free energies of both the acid molecule and its corresponding anion in a given solvent. The acid or the anion is placed in a void of a continuum medium with a dielectric constant representing the solvent.[7−9] Such a calculation is involved and prohibitively expensive for large acid molecules, and yet large uncertainties are expected in the result, mainly because the method neglects local acid−solvent interactions. Alternatively, a large enough set of solvent molecules could explicitly be included to describe acid−solvent structure and interactions.[10−12] This would further complicate the calculation and make it even more expensive, particularly for large acid molecules.

It is known that acid−base interactions result in significant changes to molecular structure and properties on both the acid and the base molecules. In the gas phase, an acid and a base molecule join together and form a hydrogen-bonded complex in which the acid acts as a hydrogen-bond donor and the base as an acceptor. Although an actual proton transfer from the acid to the base may not take place in the absence of polar or polarizable solvent molecules, the covalent bond holding the proton to the acid is significantly weakened under the influence of the base molecule.[13−17] The immediate effects include the elongation of the covalent bond holding the proton and a red-shift in the harmonic frequency for the stretch vibration of the

---

* Corresponding author. E-mail: ftao@fullerton.edu.

[†] Current address: Department of Biological Science, Stanford University, Palo Alto, CA 94305.

[‡] Analytical Chemistry, Molecular Structure, Amgen Inc., Thousand Oaks, CA 91320.

10.1021/jp710291c CCC: $40.75

bond, all compared to those for the isolated acid molecule.[13−17] The magnitude of change in these molecular properties is anticipated to reflect the acidity of a specific acid. As a result, molecular properties resulting from the formation of an acid−base hydrogen-bonded complex are to correlate with the acidity of the acid. If such correlations are established and calibrated systematically with the use of a given base, which serves as a probing molecule, the acidity of other unknown acids of similar molecular structure can be determined by the similarly calculated molecular properties of the acid complexes with the same probing molecule.

Some of the correlations have been noted in recent studies. Studies by Abkowicz-Bienko et al.[18] found that the acidities of phenols could correlate with several calculated molecular properties of their complexes with ammonia, namely, their OH bond lengths and frequencies, their hydrogen-bond lengths, and their deprotonation energies. Abraham and Platts[19] showed that hydrogen-bond structural constants of certain functional groups could relate to the compound's acid and base equilibrium constants, which have been shown to run either parallel or antiparallel to the actual acidity and basicity.[20] Previous studies have demonstrated the feasibility of using these relationships to predict acidity in chlorophenols, fluorophenols, and bromophenols[21,22] using complexes with ammonia or water. In all cases, the bond length and stretch frequency of the OH bond as well as the bond length and binding energy of the hydrogen bond were shown to be linearly correlated with available experimental pK_a values. The predicted pK_a values based on derived linear correlations from various properties were found to be accurate as compared to the experimental pK_a values.

Carboxylic acids were chosen as an important target of study for this technique, due to their simplicity, variety, and ubiquity. Carboxylic acids are generally weak acids and are reasonably stable in the protonated form, making it easy to consistently calculate the properties associated with the acidic OH bond. The acidity of a carboxylic acid is significantly affected by the presence of functional groups near the carboxyl group, allowing for a wide range of acidity to be studied. Also, carboxylic acids are the main acidic functional group found in biological systems and play important roles as biological fuels and in enzymes. For example, in the digestive enzyme trypsin, specificity is provided by a space within the structure containing an acidic aspartate residue, which ensures that only amino acids with basic side chains are cleaved.[23,24] A convenient method of determining the acidity of a single carboxyl group in a system might find significance in a wide variety of applications, such as the screening of drugs that inhibit a specific enzyme.

In the present study, carboxylic acid−ammonia complexes with known pK_a values are characterized by a set of selected molecular properties as determined by density functional theory (DFT) calculations. The properties are OH bond length, OH bond frequency, hydrogen-bond length, and hydrogen-bond energy. Correlations among these molecular properties are explored, followed by their correlations with the known pK_a values. The resulting correlations are used to predict pK_a values independently, which are compared with the known pK_a values. Disadvantages and recommendations regarding the proposed approach are discussed.

2. Methods

Two series of relatively simple carboxylic acids, aliphatic carboxylic acids and aromatic carboxylic acids (benzoic), shown in the left columns of Tables 1 and 2, respectively, are chosen to cover a large variety of substitution effects on parent acids and a wide range of pK_a values (ca. 0.5−5.0). Separate correlations for the two series of acids have turned out to be more favorable than a united treatment, probably because the carboxylic group in a benzoic acid is conjugated with the large aromatic ring and, therefore, behaves differently from a simple aliphatic carboxylic acid without such a conjugation. The separate correlations for the two series of acids may also be attributed to considerably different effects of solvent on the two groups of acids. However, similar trends are expected between the aliphatic and aromatic acid series. Since the focus of this study is primarily on the acidity from the carboxyl group, excessively large molecules with multiple polar functional

TABLE 1: Calculated Molecular Properties for Aliphatic Carboxylic Acid−Ammonia Complexes[a]

molecule	r(O−H)	ν(O−H)	r(OH···N)	ΔE_{hb}
propanoic acid	1.0099	3025.6	1.7613	−11.99
2,2-dimethylpropanoic acid	1.0100	3022.0	1.7584	−11.91
butanoic acid	1.0103	3020.6	1.7600	−9.12
pentanoic acid	1.0103	3015.9	1.7591	−11.92
2-methylbutanoic acid	1.0103	3021.7	1.7620	−11.92
acetic acid	1.0111	3008.4	1.7553	−12.14
2-methylpropanoic acid	1.0112	3009.1	1.7590	−12.02
4-hydroxybutanoic acid	1.0112	2986.2	1.7544	−12.96
3-hydroxypropanoic acid	1.0117	2992.6	1.7492	−12.25
4-chlorobutanoic acid	1.0134	2962.0	1.7409	−12.64
2-chlorobenzeneacetic acid	1.0138	2959.1	1.7392	−12.45
4-chlorobenzeneacetic acid	1.0144	2946.9	1.7365	−12.75
3-chlorobenzeneacetic acid	1.0147	2942.6	1.7345	−12.76
3-chloropropanoic acid	1.0155	2926.6	1.7298	−12.98
phenoxyacetic acid	1.0158	2919.7	1.7265	−13.35
formic acid	1.0165	2923.9	1.7341	−13.06
2-chloropropanoic acid	1.0173	2897.1	1.7194	−13.35
bromoacetic acid	1.0191	2863.4	1.7115	−14.05
chloroacetic acid	1.0193	2860.9	1.7124	−13.75
cyanoacetic acid	1.0218	2817.8	1.6983	−14.48
dichloroacetic acid-a	1.0254	2758.5	1.6829	−15.73
nitroacetic acid	1.0264	2739.4	1.6780	−15.21
trichloroacetic acid	1.0287	2694.8	1.6631	−15.46
trifluroacetic acid	1.0303	2676.1	1.6596	−15.84

[a] Bond distances r(O−H) and r(OH···N) are in angstrom (Å), harmonic frequency ν(O−H) in cm^{-1}, and hydrogen-bond energy ΔE_{hb} in kcal/mol.

TABLE 2: Calculated Molecular Properties for Substituted Benzoic Acid−Ammonia Complexes

molecule	r(O−H)	ν(O−H)	r(OH···N)	ΔE_{hb}
2-methylbenzoic acid	1.0096	3022.3	1.7540	−11.93
4-methoxybenzoic acid	1.0101	3022.3	1.7537	−11.99
4-hydroxybenzoic acid	1.0101	3022.0	1.7546	−12.16
3-methylbenzoic acid	1.0104	3006.7	1.7508	−12.19
3,5-dimethylbenzoic acid	1.0106	3006.1	1.7508	−12.10
4-methylbenzoic acid	1.0109	3000.0	1.7482	−12.16
3-methoxybenzoic acid	1.0113	2992.5	1.7464	−12.26
benzoic acid	1.0119	2984.7	1.7434	−12.39
3-hydroxybenzoic acid	1.0125	2971.6	1.7394	−12.44
2-fluorobenzoic acid	1.0127	2970.9	1.7392	−12.58
4-fluorobenzoic acid	1.0130	2963.1	1.7377	−12.69
2-chlorobenzoic acid	1.0131	2962.1	1.7367	−12.62
2-bromobenzoic acid	1.0131	2959.5	1.7353	−12.87
4-chlorobenzoic acid	1.0138	2950.3	1.7342	−12.78
4-bromobenzoic acid	1.0139	2948.0	1.7337	−12.98
3-fluorobenzoic acid	1.0145	2937.2	1.7294	−12.90
3-chlorobenzoic acid	1.0147	2934.8	1.7285	−12.94
3-bromobenzoic acid	1.0147	2934.6	1.7304	−13.18
3-formylbenzoic acid	1.0156	2918.8	1.7239	−13.13
4-formylbenzoic acid	1.0159	2912.4	1.7221	−13.17
4-cyanobenzoic acid	1.0174	2887.8	1.7151	−13.48
3-cyanobenzoic acid	1.0175	2885.2	1.7153	−13.52
3-nitrobenzoic acid	1.0183	2870.9	1.7113	−13.70
4-nitrobenzoic acid	1.0184	2868.6	1.7095	−13.68
2-nitrobenzoic acid	1.0209	2831.1	1.6977	−14.73

groups are undesirable as the experimental pK_a values may not accurately reflect the local acidity at the given carboxyl group. Moreover, compounds with bulky substituents near the carboxyl group should be treated separately with special care, due to possible steric hindrance with the probing molecule.

The syn conformer of each carboxylic acid was used due to the stability resulting from intramolecular hydrogen bonding. The carboxylic group is planar, containing an sp^2 hybridized carbon bound to two oxygen atoms and an acidic hydrogen bound to one of the oxygen atoms in the same plane. Ammonia, as a simple probing base, was placed collinearly with the acidic OH bond to form a hydrogen-bonded complex with the carboxylic acid. Such an initial geometry was subsequently optimized by full geometry optimization. The planar symmetry of the carboxylic group was maintained in the optimized hydrogen-bond structure with ammonia. Previous studies with chlorophenol−ammonia and chlorophenol−water complexes have shown that ammonia serves as a better probe than water because the stronger basicity of ammonia allows the probe to be more sensitive to acidity.[21,22]

In the case of an aromatic carboxylic acid (benzoic acid), the carboxyl group was maintained to be coplanar with the aromatic ring and hydrogen-bonded with the ammonia probe, just as in the carboxylic acids calculation. For the benzoic acid with a substitution at the ortho position, the conformation with the carboxylic OH−ammonia unit on the opposite side from the substitution was used to minimize steric hindrance by bulky substitution such as with bromine. Full geometry optimization was carried out for the hydrogen-bonded complexes using density functional theory (DFT) with Becke's three-parameter functional with the gradient-corrected correlation of Lee, Yang, and Parr (B3LYP) along with the 6-311+G(d) basis set.[25]

Geometrical parameters and molecular properties of the hydrogen-bonded complexes obtained from the DFT calculations were examined for the anticipated mutual dependence or correlations. These parameters or properties include the OH bond length, *r*(O−H); OH bond stretching frequency, *υ*(O−H); hydrogen-bond length, *r*(OH···N); and hydrogen-bond energy, ΔE_{hb}. All four properties have been shown to reflect the OH bond strength and correlate well with pK_a in chloro- and bromophenols.[21,22] A longer *r*(O−H) indicates a stronger pull on the hydrogen by the probe and thus a greater tendency toward dissociation. Likewise, a lower *υ*(O−H) indicates a weaker OH bond due to the stronger pull. The hydrogen bond to the ammonia probe is measured by *r*(OH···N) and ΔE_{hb}. A shorter *r*(OH···N) represents a stronger pull by the probe and so does the larger absolute value ΔE_{hb}; both directly indicate the stronger hydrogen bond formed by the more acidic acid and the given probe. The basis set superposition error (BSSE) effect was examined for several representative systems. The BSSE effect on ΔE_{hb} was small (~0.8 kcal/mol) and nearly constant for the different systems and was thus determined to be insignificant. The BSSE effect on molecular structure, such as *r*(O−H) and *r*(OH···N), was found to be even less important.

The calculated molecular properties were plotted against available literature pK_a values[26] and a least-squares linear regression was performed on each data set of a given property. The square of the correlation coefficient, R^2, from the regression calculations was reported to show the level of the correlation and thus the reliability of the proposed procedure in which the acidity can be predicted from the correlation. The linear relationships so established were used to recalculate pK_a. The recalculated/predicted pK_a values using each of the four different properties were evaluated for consistency and also compared with the literature values. Consensus among the predicted pK_a values was measured in terms of standard deviations from the mean, and the difference between the predicted mean and literature values was also examined to understand the quality and reliability of the proposed methods.

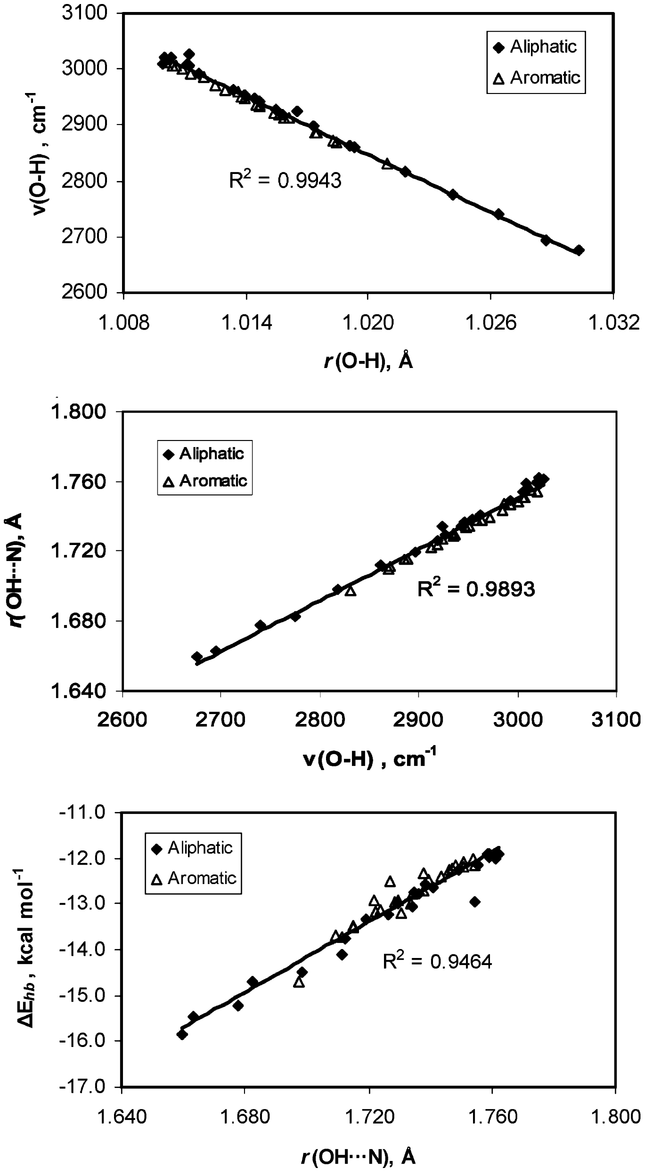

Figure 1. Correlations between molecular properties for carboxylic−ammonia complexes: (a) *r*(O−H) and *υ*(O−H), (b) *υ*(O−H) and *r*(OH···N), (c) *r*(OH···N) and ΔE_{hb}. The straight line shows the linear fit with the R^2 value for the data points.

3. Results and Discussion

3.1. Molecular Properties of Carboxylic Acid−Ammonia Complexes. Tables 1 and 2 give the OH bond lengths *r*(O−H), in ascending order, and the corresponding OH stretching frequencies *υ*(O−H), hydrogen-bond lengths *r*(OH···N), and hydrogen-bond energies ΔE_{hb} of the hydrogen-bonded complexes with ammonia for the two series of carboxylic acids, aliphatic and aromatic (benzoic), respectively. The *r*(O−H) varies from 1.010 Å for propanoic acid to 1.030 Å for trifluroacetic acid in the aliphatic series and from 1.010 Å for 2-methylbenzoic acid to 1.021 Å for 2-nitrobenzoic acid in the aromatic series (extra decimal places for the property values were shown in the two tables). The OH bond length of

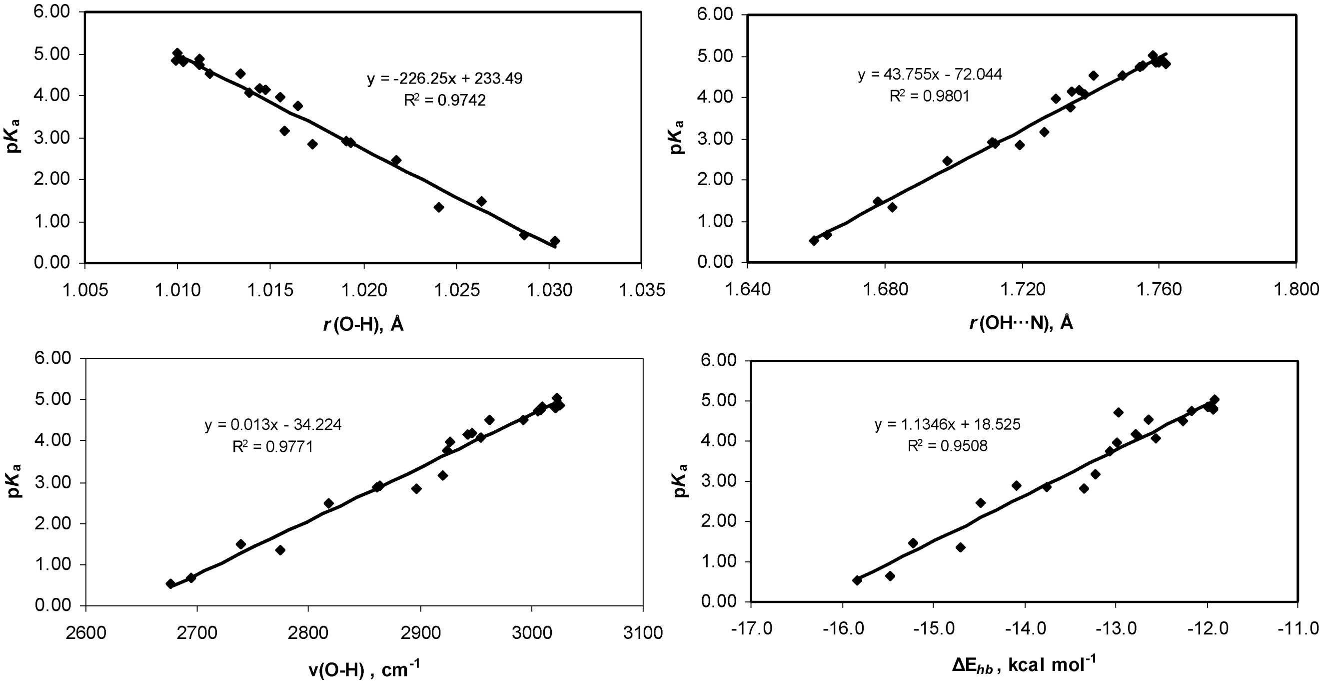

Figure 2. Correlations between pK_a and molecular properties of aliphatic carboxylic−ammonia complexes: (a) r(O−H), (b) υ(O−H), (c) r(OH···N), and (d) ΔE_{hb}. The straight line shows the linear fit with the equation and R^2 value for the data points.

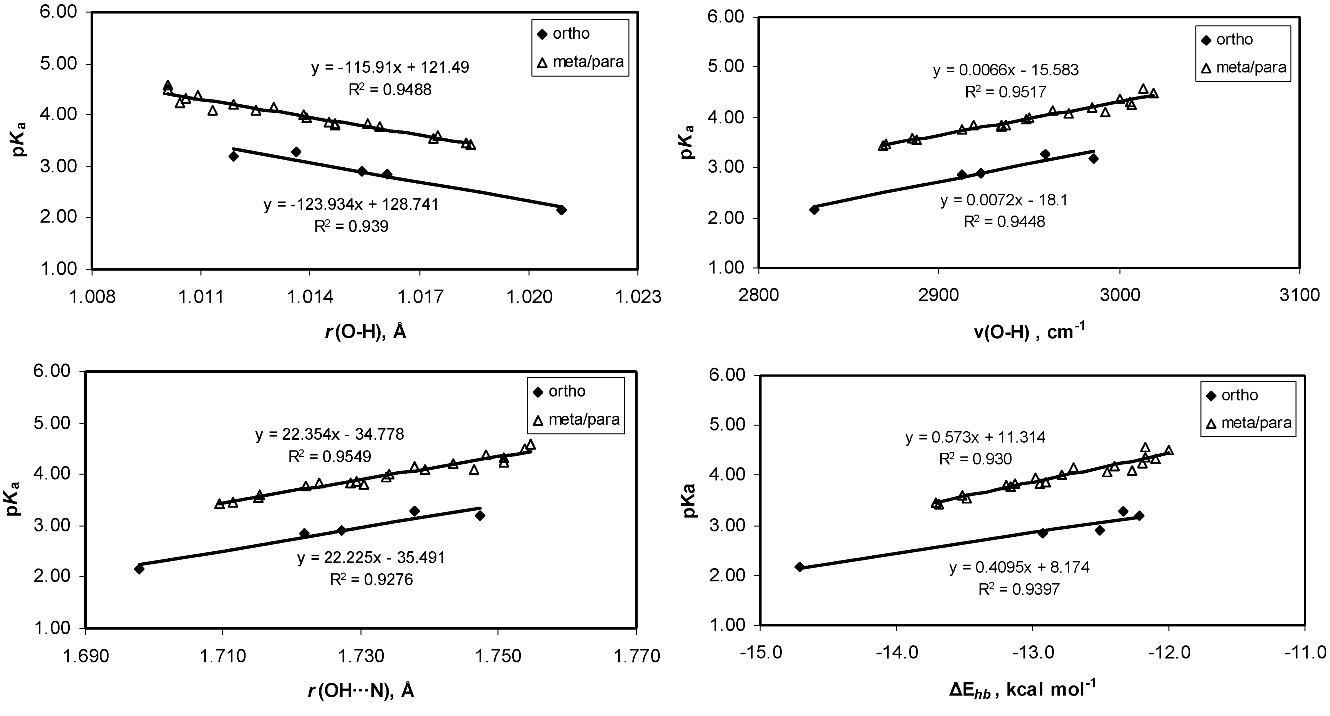

Figure 3. Correlations between pK_a and molecular properties of aromatic carboxylic−ammonia complexes: (a) r(O−H), (b) υ(O−H), (c) r(OH···N), and (d) ΔE_{hb}. The straight line shows the linear fit with the equation and R^2 value for the data points.

the carboxylic acid in the hydrogen-bonded complex with ammonia follows reasonably well the trend in the electron-withdrawing or -donating effect of the substitution group on the acid.

As the parent compounds for the two series, acetic acid and benzoic acid have r(O−H) = 1.011 and 1.012 Å, respectively. A methyl substitution, or any other aliphatic substitution, such as in butanoic acid or methylbenzoic acid, results in a shorter OH bond length than that for acidic acid or benzoic acid. Conversely, a substitution by halogen, nitro, or any other electron-withdrawing group results in a longer OH bond length than that for acetic acid or benzoic acid. The r(O−H) reflects closely the electron-withdrawing effect of the substitution group on the acid.

As shown in Tables 1 and 2, the other three properties, υ(O−H), r(OH···N), and ΔE_{hb}, just as r(O−H), also vary systematically and follow closely with the electron-withdrawing effect of the substitution group on the acid. The υ(O−H) decreases from 3025.6 cm^{-1} for propanoic acid to 2676.1 cm^{-1} for trifluroacetic acid and from 3022.3 cm^{-1} for 2-methylbenzoic acid to 2831.1 cm^{-1} for 2-nitrobenzoic acid. The r(OH···N) also decreases from 1.761 Å for propanoic acid to 1.660 Å for trifluroacetic acid and from 1.754 Å for 2-methylbenzoic acid to 1.698 Å for 2-nitrobenzoic acid. The magnitude in ΔE_{hb} increases from 11.99 kcal/mol for propanoic acid to 15.84 kcal/mol for trifluroacetic acid and from 11.93 kcal/mol for 2-methylbenzoic acid to 14.73 kcal/mol for 2-nitrobenzoic acid. The values of ν(O−H), r(OH···N), and ΔE_{hb} for acetic acid are 3008.4 cm^{-1}, 1.755 Å, and −12.14 kcal/mol and for benzoic acid are 2984.7 cm^{-1}, 1.743 Å, and −12.39 kcal/mol, respectively. A substitution by methyl or other aliphatic group leads to a higher ν(O−H), greater r(OH···N), and lower $-\Delta E_{hb}$ than the corresponding properties for acetic acid or benzoic acid. Conversely, a substitution by halogen, nitro, or any other electron-withdrawing group results in a lower ν(O−H), smaller r(OH···N), and higher $-\Delta E_{hb}$ than the corresponding properties for acetic acid or benzoic acid. It is, therefore, expected that the four properties should be mutually correlated with one another. Figure 1 shows such correlations: (a) r(O−H) and ν(O−H), (b) ν(O−H) and r(OH···N), and (c) r(OH···N) and ΔE_{hb}. The linear fit to each pair of the data in the figure is given by the straight line along with the R^2 value. It is clear that strong linear correlations exist among the four properties, r(O−H), ν(O−H), r(OH···N), and ΔE_{hb}, as shown by the high R^2 values (0.95−0.99), although the correlation with ΔE_{hb} is relatively weaker ($R^2 = 0.95$).

A variety of factors have been known to contribute to the deprotonation ability of a carboxyl group, namely, the number of substitutions and the electron-withdrawing capability, as well as the position of each substitution relative to the carboxylic group. The presence of an electron-withdrawing group, such as chlorine, bromine, or the cyano group, shifts electron density away from the proton and results in a more acidic compound. Conversely, the presence of an electron-donating group, such as the methyl or another aliphatic group, enhances electron density at the proton and hence contributes to a less acidic compound. Importantly, the magnitude of the substitution effect on acidity is sensitive to the position of substitution. No matter how strong the electron-withdrawing or electron-donating potential is for a given substitution, the effect on acidity is dramatically weakened if bound to a position that is farther from the carboxyl group.

For the two series of aliphatic acids, similar trends in substitution effects can be observed on the four properties. For the series of aromatic acids, however, the resonance effect appears to play an important role. For example, a substitution at the meta position has a nearly equal effect as one at the ortho position, although the meta position is farther away from the carboxylic group. The para substitution produces the least effect among the three positions of substitution. The similar effect from ortho and meta substitutions is likely to be a result of resonance effects.

For substituted benzoic acids, steric hindrance between the substitution group and the probing molecule may also have an effect for the ortho substitution, depending on the size of the substituent. It can be illustrated by comparing the r(O−H) between fluorobenzoic acids and cyanobenzoic acids. For the fluorobenzoic acids, the r(O−H) of the meta stereoisomer, 1.0145 Å, is greatest, followed by that of the para and ortho stereoisomers, 1.0130 and 1.0127 Å, respectively. The difference in r(O−H) between the ortho and para substitutions is considerably diminished for chloro or bromobenzoic acids, as shown by the corresponding r(O−H) values, 1.0131 and 1.0138 Å, of the stereoisomers. Such a difference in r(O−H) is more enhanced for benzoic acids with bulkier substituents. For example, the r(O−H) of *o*-cyanobenzoic acid (not included in the present study), 1.0179 Å, exceeds those of the meta and para forms, 1.0175 and 1.0174 Å, respectively. With a hydroxyl substitution, the difference is also magnified because the hydroxyl group might interact with the probing molecule.

TABLE 3: Linear Correlations of pK_a with the Different Molecular Properties for Carboxylic Acid−Ammonia Complexes

parameter	N^a	equations	R^2	SD
		Aliphatic Carboxylic Acid−Ammonia Complexes		
r(O−H)	24	pK_a = -226.25x + 233.49	0.97	0.23
ν(O−H)	24	pK_a = 0.013x − 34.224	0.98	0.22
r(OH···N)	24	pK_a = 43.755x − 72.044	0.98	0.20
ΔE_{hb}	24	pK_a = 1.1346x + 18.525	0.95	0.32
		Aromatic Carboxylic Acid−Ammonia Complexes		
r(O−H)	20 (meta/para)	pK_a = -115.91x + 121.49	0.95	0.08
	5 (ortho)	pK_a = -123.93x + 128.74	0.94	0.12
ν(O−H)	20 (meta/para)	pK_a = 0.0066x − 15.583	0.95	0.17
	5 (ortho)	pK_a = 0.0072x − 18.1	0.94	0.12
r(OH···N)	20 (meta/para)	pK_a = 22.354x − 34.778	0.95	0.07
	5 (ortho)	pK_a = 22.225x − 35.491	0.93	0.13
ΔE_{hb}	20 (meta/para)	pK_a = 0.5729x + 11.3145	0.93	0.09
	5 (ortho)	pK_a = 0.4095x + 8.174	0.94	0.12

[a] Number of compounds used for the linear correlation.

The variations in r(O−H) for the substituted benzoic acids noticed above are also present in the other properties, ν(O−H), r(OH···N), and ΔE_{hb}. These results are in contrast with those of the previous studies on halogenated phenols.[21,22] For halogenated phenols, a closer substitution consistently has a greater effect without regard to resonance.

3.2. Linear Correlations of Molecular Properties with pK_a Values. The above discussions have established that the chosen properties, r(O−H), ν(O−H), r(OH···N), and ΔE_{hb}, of the carboxylic acid complexes are all consistent with one another and with the known factors that affect the acidity. These properties reflect the relative strengths of the OH bond, as well as that of the hydrogen bond in the complex, and hence the ability of acid dissociation. To use them as predictors of acidity, however, it is important to show that these properties are quantitatively correlated to the acidity constant pK_a. The acidity constant is measured for the acid in aqueous solution, while the properties presented are calculated for the hydrogen-bonded complex of the acid in the gas phase. Solvent effects on the acid molecule are rather complicated at the molecular level and are different from one acid to another. Hence, rather complicated relationships of acidity are anticipated with respect to the molecular properties of the gas-phase hydrogen-bonded complex formed by the acid and base molecules. Figures 2 and 3 plot the available literature pK_a values for the aliphatic and aromatic carboxylic acids, respectively, against each of the four properties, r(O−H), ν(O−H), r(OH···N), and ΔE_{hb}. The straight lines are least-squares linear regression fits between the known pK_a and each property. The correlation equations from the linear regressions are shown in Table 3 along with the R^2 values and standard deviations (SD). For the aliphatic carboxylic acids, the least-squares linear regression gives high R^2 values of 0.97, 0.98, 0.98, and 0.95 for the properties r(O−H), ν(O−H), r(OH···N), and ΔE_{hb}, respectively. Clearly, all are found to be linear and strongly correlated. The corresponding SD values (in pK_a unit), 0.23, 0.22, 0.20, and 0.32, also indicate that the linear fits are of high quality. It is quite surprising to see such strong and yet simple correlations of the properties with the pK_a values.

Strong linear correlations are also found for the aromatic carboxylic acids, with the R^2 values from 0.93 to 0.95. However, two separate relationships are apparently necessary, one for those with ortho substitutions and the other for those without ortho

TABLE 4: Predicted pK_a Values for Aliphatic Carboxylic Acids from Linear Correlations

		predicted pK_a						
molecule	expt pK_a	r(O−H)	ν(O−H)	r(OH···N)	ΔE_{hb}	av pK_a	SD	diff[a]
2,2-dimethylpropanoic acid	5.03	4.98	5.06	4.89	5.00	4.98	0.07	0.05
propanoic acid	4.87	4.71	5.11	5.02	4.91	4.94	0.17	0.07
2-methypropanoic	4.84	5.00	4.89	4.92	4.93	4.94	0.04	0.10
butanoic acid	4.83	4.91	5.04	4.96	4.97	4.97	0.06	0.14
pentanoic acid	4.83	4.91	5.04	4.93	5.00	4.97	0.06	0.14
2-methylbutanoic acid	4.80	4.91	5.06	5.05	5.00	5.00	0.07	0.20
acetic acid	4.76	4.73	4.89	4.76	4.72	4.77	0.08	0.01
4-hydroxybutanoic acid	4.72	4.71	4.85	4.72	3.82	4.52	0.48	0.20
4-chlorobutanoic acid	4.52	4.21	4.28	4.13	4.19	4.20	0.06	0.32
3-hydroxypropanoic acid	4.51	4.59	4.68	4.49	4.61	4.59	0.08	0.08
4-chlorobenzeneacetic acid	4.19	3.98	4.09	3.94	4.03	4.01	0.06	0.18
3-chlorobenzeneacetic acid	4.14	3.91	4.03	3.85	4.05	3.96	0.10	0.18
2-chlorobenzeneacetic acid	4.07	4.10	4.18	4.02	4.27	4.14	0.11	0.07
3-chloropropanoic acid	3.98	3.73	3.82	3.64	3.80	3.75	0.08	0.23
formic acid	3.75	3.51	3.79	3.83	3.71	3.71	0.14	0.04
phenoxyacetic acid	3.17	3.67	3.73	3.50	3.52	3.61	0.11	0.44
bromoacetic acid	2.90	2.92	3.00	2.84	2.53	2.82	0.20	0.08
chloroacetic acid	2.87	2.87	2.97	2.88	2.92	2.91	0.04	0.04
2-chloropropanoic acid	2.83	3.33	3.44	3.19	3.37	3.33	0.11	0.50
cyanoacetic acid	2.47	2.31	2.41	2.27	2.09	2.27	0.13	0.20
nitroacetic acid	1.48	1.27	1.39	1.38	1.25	1.32	0.07	0.16
dichloroacetic acid	1.35	1.79	1.85	1.57	1.85	1.76	0.14	0.41
trichloroacetic acid	0.66	0.75	0.81	0.72	0.98	0.81	0.11	0.15
trifluoroacetic acid	0.52	0.38	0.57	0.57	0.56	0.52	0.09	0.00

[a] Calculated as the absolute difference of experimental and average predicted pK_a.

substitutions. In the previous studies on halogenated phenols, separate correlations were necessary for those with double ortho substitutions. This was found to be the result of steric interference from the ortho substituent, which caused the ammonia probe to be pushed away from a position for a linear hydrogen bond and thus produce a different correlation. It was initially thought that the similar effect might be responsible for the separate correlations of the aromatic carboxylic acids with ortho substitutions. A closer examination reveals that any ortho substitution appears to increase the acidity of the acid, regardless of the electric effect of the ortho substitution on the benzoic acid. Even an electron-donating group would raise the acidity considerably from that of the parent benzoic acid. For example, 2-methylbenzoic acid has a pK_a of 3.19, which is compared to 4.20 for benzoic acid.

The abnormal behavior, however, is not shown in the molecular properties of the ortho-substituted benzoic acids. In fact, 2-methylbenzoic acid has the perfectly anticipated values of r(O−H), υ(O−H), r(OH···N), and ΔE_{hb}, which would make the acid corresponding to the weakest among all of the substituted benzoic acids considered in the present study. Instead, 2-methylbenzoic acid is not only stronger than the parent benzoic acid but it is also stronger than all of the non-ortho-substituted benzoic acids, most of which have strong electron-withdrawing groups. What is the true origin in the unusually strong acidity of all ortho-substituted benzoic acids, which also merits totally separate correlations of pK_a with r(O−H), υ(O−H), r(OH···N), and ΔE_{hb}? No clear answer is offered from the present study, but it is likely a result of unique interactions between the ortho subsituent and the carboxylic group.

3.3. Prediction of pK_a Values from Molecular Properties. The strong correlations established for the pK_a with each of the four properties, r(O−H), υ(O−H), r(OH···N), and ΔE_{hb}, should allow the prediction of the pK_a for any carboxylic acid similar in structure to those included in obtaining the correlations. To test the performance of such an approach, the pK_a values of the acids considered in this study were recalculated using each of the correlations. Although such "predictions" are nothing but a closer inspection of the linear fits against the literature pK_a's on individual acids, the results may help better understand how well the proposed strategy would perform in practical applications. Tables 4 and 5 present the predicted pK_a values from the correlation equations (Table 3) for the aliphatic and aromatic acids, respectively. Figure 4 plots the literature pK_a verses the average of the predicted pK_a values from the four properties.

As shown in Tables 4 and 5, the predicted pK_a values from the four properties are surprisingly consistent. The standard deviations from the average are within 0.1 unit in pK_a for the majority of the acids. The largest SD, 0.48, occurs for 4-hydroxybutanoic acid, which appears to be caused by an inconsistent prediction of the pK_a using the ΔE_{hb} correlation. The second largest SD, 0.20, happens for bromoacetic acid and is also attributed to the poor prediction using the ΔE_{hb} correlation. As shown in Figure 4, the average values of the predicted pK_a's are in good agreement with the literature values, and their correlations have $R^2 = 0.98$ for both of the aliphatic and aromatic series. However, there are exceptionally large deviations (Table 4 "diff" column) from the literature values for several acids in the aliphatic species, such as dichloroacetic acid (0.41), phenoxyacetic acid (0.44), and 2-chloropropanoic acid (0.50). These large deviations indicate that the literature values of these acids may not be as reliable as those of other acids in the series. There could be other causes for the outliers too. Phenoxyacetic acid, for example, has an aromatic ring, which could make it different from other aliphatic acids in the series.

Table 6 summarizes the overall rms deviations between the calculated and experimental pK_a values. Among the four properties, r(O−H), υ(O−H), r(OH···N), and ΔE_{hb}, used for the prediction of pK_a, all three except for ΔE_{hb} are capable of providing exceptionally accurate and consistent predictions. The prediction using ΔE_{hb} is less accurate and results in large deviations for several acids, perhaps due to influences from a variety of other factors. However, the prediction using r(O−H) or r(OH···N) stands out to be the most convenient and yet capable of giving the top performance and is therefore recom-

TABLE 5: Predicted pK_a Values for Aromatic Carboxylic Acids from Linear Correlations

		predicted pK_a						
molecule	expt pK_a	r(O−H)	ν(O−H)	r(OH···N)	ΔE_{hb}	av pK_a	SD	diff[a]
4-hydroxybenzoic acid	4.57	4.41	4.30	4.44	4.34	4.37	0.06	0.20
4-methoxybenzoic acid	4.50	4.41	4.34	4.42	4.44	4.40	0.04	0.10
4-methylbenzoic acid	4.37	4.32	4.22	4.30	4.34	4.29	0.05	0.08
3,5-dimethylbenzoic acid	4.32	4.35	4.26	4.36	4.38	4.34	0.06	0.02
3-methylbenzoic acid	4.25	4.37	4.26	4.36	4.33	4.33	0.05	0.08
benzoic acid	4.20	4.20	4.12	4.19	4.21	4.18	0.04	0.02
4-fluorobenzoic acid	4.15	4.07	3.97	4.07	4.04	4.04	0.05	0.11
3-methoxybenzoic acid	4.10	4.27	4.17	4.26	4.29	4.25	0.05	0.15
3-hydroxybenzoic acid	4.08	4.13	4.03	4.10	4.18	4.11	0.06	0.03
4-chlorobenzoic acid	4.00	3.98	3.89	3.99	3.99	3.96	0.05	0.04
4-bromobenzoic acid	3.96	3.97	3.87	3.98	3.88	3.92	0.06	0.04
3-fluorobenzoic acid	3.86	3.90	3.80	3.88	3.92	3.88	0.05	0.02
3-chlorobenzoic acid	3.84	3.88	3.79	3.86	3.90	3.86	0.05	0.02
3-formylbenzoic acid	3.84	3.77	3.68	3.76	3.79	3.75	0.05	0.09
3-bromobenzoic acid	3.81	3.88	3.79	3.90	3.76	3.83	0.07	0.02
4-formylbenzoic acid	3.77	3.74	3.64	3.72	3.77	3.72	0.06	0.05
3-cyanobenzoic acid	3.60	3.55	3.46	3.57	3.57	3.54	0.05	0.06
4-cyanobenzoic acid	3.55	3.56	3.48	3.56	3.59	3.55	0.05	0.00
3-nitrobenzoic acid	3.46	3.46	3.36	3.48	3.46	3.44	0.05	0.02
4-nitrobenzoic acid	3.43	3.45	3.35	3.44	3.47	3.43	0.05	0.00
2-fluorobenzoic acid	3.27	3.12	3.21	3.13	3.12	3.15	0.04	0.12
2-methylbenzoic acid	3.19	3.34	3.40	3.34	3.17	3.31	0.10	0.12
2-chlorobenzoic acid	2.90	2.90	2.95	2.90	3.05	2.95	0.07	0.05
2-bromobenzoic acid	2.85	2.81	2.87	2.77	2.88	2.83	0.05	0.02
2-nitrobenzoic acid	2.17	2.22	2.28	2.24	2.15	2.22	0.06	0.05

[a] Calculated as the absolute difference between experimental and average predicted pK_a.

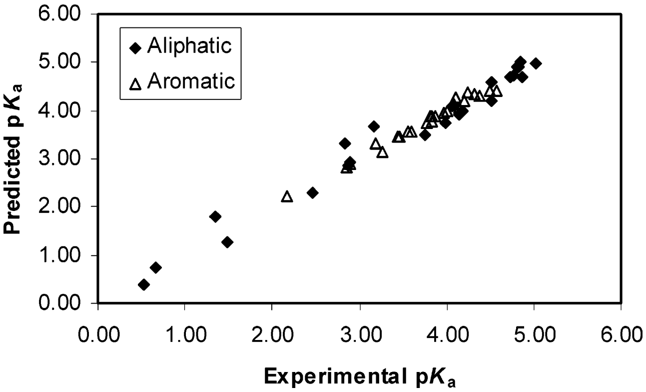

Figure 4. Plot of the mean pK_a values from the linear correlations and the experimental pK_a values for the aliphatic and aromatic carboxylic acids.

TABLE 6: Root-Mean-Square (rms) Deviations from Experimental pK_a

	r(O−H)	ν(O−H)	r(OH···N)	ΔE_{hb}
aliphatic	0.220	0.240	0.193	0.299
aromatic	0.079	0.117	0.077	0.088

mended from the present study. The prediction using ν(O−H) is equally good, but it is much more demanding in computer resources (involving the calculation of vibrational frequencies). In contrast, the prediction using r(O−H) or r(OH···N) only requires geometry optimization. This method is more attractive if only partial optimization for the active geometrical parameters is carried out with the remaining molecular geometry frozen at that of the monomer.

4. Conclusions

Molecular properties of substituted carboxylic acids and benzoic acids in the hydrogen-bonded complexes with ammonia have been calculated using density functional theory and examined for their interdependence as well as dependence on the acidity of the acid. The bond length r(O−H) and stretching frequency ν(O−H) of the hydroxyl group and the hydrogen-bond length r(OH···N) and energy ΔE_{hb} of the complex are shown to be highly correlated with one another and are linearly correlated with available literature pK_a values of the carboxylic acids. The linear correlations resulting from the fit to the available pK_a values are used to recalculate the pK_a values of all substituted carboxylic acids and benzoic acids in the series. The pK_a values so calculated from the different molecular properties are highly consistent and in good agreement with the literature values. It is suggested from this study that effective and reliable predictions of pK_a values for organic acids are feasible using the linear correlations established for a set of acids with known acidities. This approach is unique in its capability to determine the acidity of a particular functional group or the local acidity within a large molecular system, such as a protein.

The chosen molecular properties, r(O−H), ν(O−H), r(OH···N), and ΔE_{hb}, produce reliable predictions due to their fundamental connections to the acidity of the acid. The predictions from r(O−H), ν(O−H), and r(OH···N) are nearly equal in the top performance, while the prediction from ΔE_{hb} is slightly less accurate. In terms of practical applications, however, r(O−H) or r(OH···N) is recommended, due to the simplicity and economics in the computation. Separate correlations are necessary for different classes of organic acids. Aliphatic and aromatic carboxylic acids are two different classes that require separate correlations. Even within the aromatic carboxylic acids, those with ortho substitutions belong to a different group and require a correlation different from that for those without ortho substitutions.

The new approach presented here and in the previous studies may have applications in many areas in which acidity is not easily determinable in a traditional way. For example, in a compound with multiple acidic sites, the acidity of one particular site can be measured using the new approach, whereas in traditional methods the acidity relates to the entire molecule in question. This would be especially useful in large, complex molecules, such as proteins, in which a particular acidic site could be treated with density functional theory calculations to find r(O−H) or r(OH···N) values and, ultimately, the acidity value using the established correlation for the appropriate class of relatively simple acids.

Acknowledgment. This work was supported in part by The Camille and Henry Dreyfus Foundation (Award No. TH-00-028), The California State University Wang Family Scholarship, and National Science Foundation of China (Grant No. 25028706).

References and Notes

(1) Bartak, P.; Cap, L. *J. Chromatogr. A* **1997**, *767*, 171.

(2) Li, X.-J.; Zeng, Z.-R.; Zhou, J.-J. *Anal. Chim. Acta* **2004**, *509*, 27.

(3) Da Silva, C. O.; Da Silva, E. C.; Nascimento, A. C. *J. Phys Chem. A* **1999**, *103*, 11194−11199.

(4) Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. *J. Phys. Chem. A.* **2002**, *106*, 1327−1335.

(5) Kaminski, G. A. *J. Phys Chem. B.* **2005**, *109*, 5884−5890.

(6) Fersht, A. *Structure and Mechanism in Protein Science*; Freeman and Co.: New York, 1999; pp 169−190.

(7) Onsager, L. *J. Am. Chem. Soc*. **1938**, *58*, 1486−1493.

(8) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys*. **1981**, *55*, 117.

(9) Modesto, O.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187−4225.

(10) Milet, A.; Struniewicz, C.; Moszynski, R.; Wormer, P. E. S. *J. Chem. Phys.* **2001**, *115*, 349−356.

(11) Tachikawa, M. *Mol. Phys*. **2002**, *100*, 881−901.

(12) Knochenmuss, R.; Leutwyler, S. *J. Chem. Phys*. **1989**, *91*, 1268−1278.

(13) Nguyen, M.-T.; Jamka, A. J.; Cazar, R. A.; Tao, F.-M. *J. Chem. Phys.* **1997**, *106*, 8710−8717.

(14) Cazar, R. A.; Jamka, A. J.; Tao, F.-M. *Chem. Phys. Lett.* **1998**, *287*, 549−552.

(15) Cazar, R. A.; Jamka, A. J.; Tao, F.-M. *J. Phys. Chem*. **1998**, *102*, 5117−5123.

(16) Snyder, J. A.; Cazar, R. A.; Jamka, A. J.; Tao, F.-M. *J. Phys. Chem. A* **1999**, *103*, 7719−7724.

(17) Conley, C.; Tao, F.-M. *Chem. Phys. Lett*. **1999**, *301*, 29−36.

(18) Abkowicz-Bienko, A. J.; Latajka, Z. *J. Phys Chem. A.* **2000**, *104*, 1004−1008.

(19) Abraham, M. H.; Platts, J. A. *J. Org. Chem*. **2001**, *66*, 3484−3491.

(20) Alkorta, I.; Campillo, N.; Rozas, I.; Elguero, J. *J. Org. Chem.* **1998**, *63*, 7759−7763.

(21) Han, J.; Tao, F.-M. *J. Phys. Chem. A* **2006**, *110*, 257−263.

(22) Han, J.; Deming, R. L.; Tao, F.-M. *J. Phys. Chem. A* **2005**, *109*, 1159−1167.

(23) Graf, L.; et al. *Biochemistry* **1987**, *26*, 2616−2623.

(24) Krieger, M.; Kay, L. M.; Stroud, R. M. *J. Mol. Biol.* **1974**, *83*, 209−230.

(25) Frisch, M. J.; et al. *Gaussian 98,* Revision A.11, Gaussian, Inc.: Pittsburgh, PA, 2001.

(26) Lide, D. R. (Ed.) *CRC Handbook of Chemistry and Physics*, 77th ed.; CRC Press: Boca Raton, FL, 1996.