Theoretical Prediction of the Abraham Hydrogen Bond Acidity and Basicity Factors from a Reaction Field Method

Ivo Cacelli,*,[†] Silvio Campanile,[‡] Alessandro Giolitti,[§] and Dafne Molin[‡]

Dipartimento di Chimica e Chimica Industriale, Università di Pisa, via Risorgimento 35, 56126 Pisa, Italy,
Istituto per i Processi Chimico Fisici del CNR, Area della Ricerca, Via G. Moruzzi 1, I-56010 Pisa, Italy, and
Menarini Ricerche SpA, Via Sette Santi 3, I-50131 Firenze, Italy

A new methodology for the theoretical evaluation of the hydrogen bond acidity $\Sigma\alpha_2^H$ and basicity $\Sigma\beta_2^H$ Abraham descriptors is presented. The first step is a quantum mechanical calculation at the Hartree−Fock level using a moderate basis set, including the solute−solvent interaction through a Reaction Field method, namely the Polarizable Continuum Model (PCM). The density charge on the surface of the cavity surrounding the solute, which contains the signature of the specificity of the molecule, is then translated into effective atomic charges through a suitable algorithm. These atomic charges can be related to the acidity and basicity properties of the molecule by a proper parametrization of empirical atomic factors, which account for the specific H-bonding capabilities of the individual atoms and group of atoms. The Abraham descriptors can be then evaluated with a high degree of accuracy for a large number of classes of molecules. Calculations performed for a set of 55 compounds give a standard deviation of 0.029 and 0.044 for $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$, respectively. The correlation coefficients are 0.994 and 0.974.

INTRODUCTION

Phenomena-like diffusion through membranes,[1] partition between liquid−liquid[2] or liquid−gas[3] phases, adsorption[4] or simple chemical reactions[5] include very complicated and specific interactions between molecules or group of molecules. Their comprehension would require in principle the study of the bulk properties of solutions, crystals and membranes which might be performed only at the price of a strongly simplified model with a small predictive ability. The need of simplifying these phenomena and, particularly, to avoid bulk phase studies, has brought the researchers to investigate the possibility of describing complicated problems by recurring to individual properties of the relevant chemical species. This is of course a drastic simplification as the specific interaction of two or more molecules is considered as resulting in some way from nonspecific properties i.e., determined for the isolated molecule or for a simplified system with no reference to the real complex system under examination. Nevertheless the recognition of the basic mechanisms of the interaction between pairs, like hydrogen bonding or electrostatic, permits to reach accurate predictions of the behavior of a pair of molecules using some individual properties. One of the main successes on this approach is due to Abraham, who, on the basis of previous works of Taft et al.,[6] Raevsky et al.[7] and himself, was able to predict with high accuracy the equilibrium constant of the 1:1 reaction in tetrachloromethane at 298 K

$$A-H + B \rightleftharpoons A-H\cdots B \qquad (1)$$

by the hydrogen bond (HB) acidity α_2^H factor of the AH molecule and by the hydrogen bond basicity β_2^H factor of the B molecule, through the relation[8,9]

$$\text{Log } K_{AB} = -1.094 + 7.354\ \alpha_2^H\ \beta_2^H \qquad (2)$$

The accuracy of this equation was estimated by the authors through a standard deviation of 0.09. To extend this equation to solvation properties, these hydrogen bond scales were generalized to any concentration of the involved molecules, and the effective $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ scales were determined. Together with other molecular descriptors, they can be effectively inserted in linear free energy relationships (LFER's) for the prediction of more complex phenomena. These $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ quantities are the object of our interest in the present paper (they will be called α and β for simplicity).

Because of some experimental difficulties in measuring accurate molecular descriptors, alternative methods have been proposed to evaluate α and/or β or other more or less equivalent molecular descriptors. Most of them start with a quantum mechanical (QM) calculation[10−13] and look for some correlation between molecular properties and experimental descriptors. Fragmentation methods[14] modeled on experimental data have also been proposed. Since an accurate and complete list of such attempts has been already reported and discussed in ref 15, we prefer not to go here into further details. It is however worthwhile mentioning some attempts to predict molecular properties using only theoretical descriptors derived from QM calculations: Klamt,[16] Famini and Wilson[17] and, recently, Platts et al.[18]

* Corresponding author phone: +39 050 221 9213; fax: +39 050 221 9260; e-mail ivo@dcci.unipi.it.
† Università di Pisa.
‡ Istituto per i Processi Chimico Fisici del CNR.
§ Menarini Ricerche SpA.

In this contribution we present a new method based on the reaction field approach which was originally conceived by Onsager[19] to compute solvation energies. The basic idea of representing the solvent as a continuum polarizable medium surrounding a single solute molecule inside a given cavity opened the way to more elaborate models for the solute solvent interactions[16,20,21] in which the cavity was modeled on the shape of the solute molecule. Despite the crude approximation of representing the solvent molecules as a classical medium sensitive to electrostatic forces, these methods have gained some popularity, essentially for their simplicity, which permits to avoid calculations of the supermolecule formed by the solute and a by number of solvent molecules. The latter possibility involves the complicated problem of numerous quasi-degenerate geometrical configurations as well as of larger computational requests. On the other hand, in the cavity based methods, only electrostatic forces can be explicitly considered, and therefore they present the serious problem of neglecting the specific solute solvent interactions and in particular hydrogen bonds. In the following we will show that this serious shortcoming can be circumvented by a suitable empirical parametrization of some atomic parameters.

This work is conceived as the first step of a sequence of methodologies which should lead to the estimate of a number of molecular descriptors to be used for the theoretical prediction of the biological behavior for molecules of pharmacological interest. For instance, one of the prominent difficulties in optimizing a new oral drug is still to correlate its structure to oral bioavailability. It is well-known that at least hydrogen bonding properties are critical to oral absorption, so that their appropriate description in terms of structural features could steer molecular optimization. In order for this project to be of help to the pharmaceutical science, one of its basic requirements is the possibility of extending the computation to large molecules as well as to large sequences of molecules, with a moderate computational effort. This rules out the possibility of adopting complicated and time-consuming post HF methods and, moreover, forces us to employ a moderate basis set.

THEORY

The molecule under scrutiny is considered to interact with a general solvent within the Polarizable Continuum Model (PCM)[21] that is derived by the Onsager Reaction Field model.[19] In this picture the solvent is modeled by an infinite continuum medium (with dielectric constant ϵ) outside a polyhedral cavity surrounding the solute molecule. The nuclear and electronic distribution of the solute originates an electric field which polarizes the medium and induces a charge distribution σ on the cavity surface which forms the reaction field potential V acting in its turn on the solute molecule. The nonlinear Schrödinger equation to be solved includes the electrostatic Hamiltonian in a vacuum H plus the local potential V

$$[H + V(\epsilon,\rho,\Omega)]|\Psi\rangle = E|\Psi\rangle \tag{3}$$

which depends on the dielectric constant ϵ, on the shape of the cavity Ω and on the density function ρ of the ground state Ψ. Since the Hamiltonian depends on its ground-state eigensolution, this equation has to be solved in a self-consistent way.

The faces of the cavity surrounding the molecule are called "tesserae", and it is supposed that the charge density σ is constant on each tessera. The charge (or density charge) on the tesserae can be determined by the appropriate equations of electrostatic[22] at the plane separating the vacuum by a dielectric continuum medium. The choice of the cavity is crucial in obtaining reliable results, and it is in general determined in a semiempirical way, by fittings with experimental solvation energies.[23]

This approach is clearly suitable for weak solute solvent interactions. In the case of relevant overlap between the wave function of the solute and that of the solvent, the response of the solute is expected to be inadequate, and some expedients have to be used in order to drive the interaction energy or the density charge toward acceptable results.[24] Nevertheless the σ distribution (or the reaction field) contains the chemical identity of the solute and can be used as a starting point to obtain some physical quantities referred to the solute. By this point of view the continuum representation of the solvent is no more than a probe for the response properties of the solute. This approach is in line with the above-mentioned attempt of factorizing the interaction between two (or more) molecules by separate contributions. An approach based on a similar philosophy was proposed by Klamt et al.[25] in the COSMO−RS method that uses the charge distribution on the cavity surface to build the σ-potential, which is the key function to estimate molecular properties such as ΔG and *Log*P.

In the working hypothesis that the σ distribution on the cavity surface includes important information about the molecule, we look for a connection with some quantities impressed in the atoms of the molecule. We suppose that the σ distribution, which arises from the electrostatic interaction with the nuclei and electrons and includes the self-interaction contribution, can be thought of as arising from the nuclei of the molecule bearing an effective charge.

According to the electrostatic laws, the surface density charge of the i-th tessera σ_i is proportional to the component ϵ_i of the electric field normal to the surface (directed outside the cavity), generated by the solute molecule

$$\sigma_i = q_i/S_i \propto -\epsilon_i \tag{4}$$

where q_i and S_i are the total charge and, respectively, surface of the i-th tessera. The same electric field arising from the nuclei is

$$\epsilon_i = \sum_{\mu}^{atoms} Z_{\mu} \frac{\hat{R}_{\mu i} \cdot \hat{n}_i}{R_{\mu i}^2} = \sum_{\mu}^{atoms} Z_{\mu} f_{\mu i} \tag{5}$$

here n_i is the unit vector normal to the surface S_i and $R_{\mu i}$ is the distance between the atom μ and the tessera i. The Z_{μ} are the effective nuclear charges which can be found by minimizing the functional

$$I = \sum_i^{tess} [\sigma_i - \sum_{\mu} Z_{\mu} f_{\mu i}]^2 w_i / \sum_i^{tess} w_i \tag{6}$$

where the w_i are weights included for generality reasons. By imposing the electro neutrality condition through a

Lagrangian multiplier α, the wanted effective nuclear charges must obey to the linear equation system

$$\frac{\partial}{\partial Z_\nu}[I - \alpha\sum_\mu Z_\mu] = 0 \quad \sum_\mu Z_\mu = 0 \qquad (7)$$

whose solution can be achieved by standard methods.

In this way the complex σ distribution is translated into simpler quantities which retain a clear chemical meaning. Effective atomic charges have a long history in the field of molecular structure. Since they have no rigorous meaning (i.e. they are not physical observables) a number of proposals were done in the last forty years within the so-called population analysis methods. We mention only the most known variants based on the electronic density function: Mulliken,[26] Löwdin,[27] Mayer,[28] Natural Bond Orbitals[29] and Atoms in Molecules.[30] However, our atomic charges present relevant differences since they include the response of the molecule to an external stress induced by the molecule itself and appear suitable to model the general reactivity of the molecule.

Once the Z_μ are determined, the α and β factors are computed by

$$\alpha = \sum_\mu^{pos} Z_\mu\gamma_\mu \qquad (8a)$$

$$\beta = -\sum_\mu^{neg} Z_\mu\gamma_\mu \qquad (8b)$$

where the atomic factors (AF) γ are specified for each type of atom or group of atoms and include effects of the chemical environment. In eqs 8 the summation is extended over the positively (negatively) charged atoms. For instance different atomic factors are allowed for O, N, CH_3, and they can also be different for the same atom, to account for the possible different atom they are bonded to. For instance hydrogen atoms bonded to nitrogen and to oxygen can take different AF's. Dependence from sp^2/sp^3 hybridization can be also considered.

The atomic factors embody a link between the solute–solvent electrostatic interaction and the atomic electronegativity as well as the atomic capability of forming partial bond with other (solvent) atoms. In the present method they are determined by linear regression with experimental Abraham factors. Thus they are semiempirical quantities which have to be invoked in order to overcome the difficulties, inherent in the reaction field methods, in describing solute–solvent interactions characterized by some overlap between neighboring molecules.

The aim of this method is to estimate α and β factors from single molecule calculations, i.e., considering only the reactants of the reaction 1. This is of course a difficult task because, for instance, the ΔE of the reaction, which was obviously proven to correlate with the basicity factor,[12] remains unknown. It is not surprising that this lack of information has to be complemented with some specific empirical adjustments in order to account for some general characteristics of the products in eq 1.

The first expedient was suggested by Platts[13] who observed that, despite esters could in principle undergo one H-bonding on both the basic oxygen atoms, this does not happen, probably because the attack of the first proton prevents the molecule to form a second hydrogen bond. This suggests that eq 8b has to be corrected in order to prevent the unphysical feature that all the basic atoms can be independently protonated or, in other words, that the basic propensity of the molecule is given by the sum of the atomic contributions. A similar conclusion has been recently reported[31] in the attempt to compute the hydrogen bond acidity and basicity factors, starting from molecular quantities obtained by DFT calculations. Thus equations (8) are to be modified by including a sort of attenuation factor

$$\alpha = \sum_\mu^{pos} Z_\mu\gamma_\mu(1 - \exp(-t\cdot d_{\mu0})) \qquad (9a)$$

$$\beta = -\sum_\mu^{neg} Z_\mu\gamma_\mu(1 - \exp(-t\cdot d_{\mu0})) \qquad (9b)$$

where $d_{\mu0}$ is the distance between the atom μ and the most positive (negative) charged atom and conventionally d_{00} is set to infinity. The exponential factor t was empirically fixed to 0.3 Å^{-1} in order to get the best possible α and β results. According to the conventional chemical wisdom, this functional form preserves the separated effects of two atoms bearing charges of the same sign, when they are very distant. This procedure was not applied to the acid hydrogens of NH_2 since they are nearly equivalent by symmetry and from a statistical viewpoint a double probability of H-bonding with a given basis may be roughly expected.

The second correction to be implemented is concerned with the presence of aromatic rings directly connected to the most basic or acidic atom. While the inductive effect is expected to be, at least partially, accounted for the QM calculation, attention has to be paid in those cases where resonance structures have different stabilization effects on the free molecule and on the molecule which undergoes a hydrogen bond. One example is provided by phenol, whose experimental HB acidity factor of 0.60 is much greater than that of ethyl alcohol of 0.37. The usual explanation invokes a delocalization of the partial negative charge on the ring in quinoid resonance structures, which is more effective in the molecular complex and leads to a relevant increase of acidity. This effect was simply included in our model by increasing α value by 0.1 units.

A third empirical correction was implemented in the halogenated compounds to account for the different magnitude of the inductive effect in the free and hydrogen bonded molecule. Test calculations have shown that the inclusion of one or more halogen atoms in carboxylic acids or alcohols originates an insufficient inductive effect: i.e., the substitution of one hydrogen with a halogen causes the atomic charges to change too little. Thus we emphasize the inductive effect of halogens by a further correction on the most positive and negative charged atoms of the molecule, through the formula

$$Z_0(corrected) = Z_0 \prod_{i=1}^{NHal}\left(1 \pm \frac{0.16\cdot EN_i}{N_b\cdot i}\right) \qquad (10)$$

where Z_0 refers to the most charged atoms, EN_i is the electronegativity of the i-th halogen atom, *NHal* is the

Table 1. United Atom (Å) Radii Used in the PCM Calculation

CH$_3$, CH$_2$, all C's	1.80	C (=O)	1.70
O (sp^3)	1.55	F	1.70
O (sp^2)	1.60	Cl	1.80
N (amines, nitro, nitriles)	1.60	S	1.70
N (amides)	1.65	H (alcohols, amines, thiols)	1.20
H (carboxylic acids)	1.00	H (amides)	0.95

number of halogen atoms and N_b is the number of bonds between the two atoms. The sign + and − are employed respectively for the most positive and negative atoms. The result is that the acidity is increased and the basicity is depressed. The factor 0.16 appearing in eq 10 was roughly chosen to minimize the functional (6).

COMPUTATIONAL DETAILS

The internal geometry of each molecule was first optimized at the DFT-B3LYP/cc-PVDZ level with no reaction field included. These structures are subsequently used for a Hartree−Fock SCF calculation with the standard 6-31G* basis set including the solute solvent interaction through the PCM model. All structures were optimized with the Gaussian 98 package;[32] the PCM calculations were carried out with the Gamess[33] program, modified in order to prepare a file containing some details of the reaction field. The solvent surface charge distribution on the cavity of each molecule was then processed with a program coded in our group, to obtain the effective atomic charges and then the theoretical α and β values as described above.

The cavity surrounding the molecule was determined on the basis of the United Atom approach for PCM calculations[17] in which only atoms other than hydrogen bear an individual sphere; the hydrogens are included in the same sphere of the atom they are bonded to. In the original paper the radii of the spheres were optimized in order to reproduce as well as possible the solvation energy of a variety of compounds in water. In the present calculation the united atom prescriptions have been followed for the CH$_3$, CH$_2$ and CH groups, whereas for the NH, OH, and SH groups both atoms bear their own sphere (the radius of the sphere centered on the heteroatom has been consequently decreased). This change was necessary in order to increase the sensitivity of the reaction field to acid hydrogen atoms. Consistently with the united atom approach, the effective atomic charges are only allowed for the atoms or groups of atoms bearing a sphere: CH$_3$, CH$_2$, CH, O N, H, halogens. The resulting effective atomic charges and, consequently, the resulting acidity and basicity factors show little sensitivity to the value of the radii at least for changes of ± 0.1−0.2 Å. So we decided to differentiate as little as possible the radii of the same atom in the different chemical contexts and only to make few refinements to get some improvement in the final results. To give the needed information for the reproducibility of our results, a complete list of the employed radii is reported in Table 1.

A set of 55 compounds was selected from the experimental database of α and β values.[9] Molecules were chosen in order to consider the most common functional groups in organic chemistry, including halogen and sulfur atoms, and to cover a wide range of acidity and basicity. Compounds with two or more basic sites and one or two acidic hydrogen atoms were included in the sample, to test the method for a large

Table 2. Sample of Compounds Used in the Present Paper with Experimental and Computed Acidity and Basicity Factors

		α		β	
no.	name	exp	calc	exp	calc
1	acetic acid	0.61	0.62	0.44	0.46
2	propanoic acid	0.60	0.61	0.45	0.46
3	butanoic acid	0.60	0.61	0.45	0.47
4	benzoic acid	0.59	0.63	0.40	0.48
5	*p*-methylbenzoic acid	0.60	0.60	0.38	0.43
6	chloroacetic acid	0.74	0.72	0.36	0.43
7	trichloroacetic acid	0.95	0.92	0.28	0.37
8	methanol	0.43	0.39	0.47	0.27
9	ethanol	0.37	0.39	0.48	0.45
10	1-butanol	0.37	0.36	0.48	0.44
11	pentan-1-ol	0.37	0.35	0.48	0.47
12	hexan-1-ol	0.37	0.34	0.48	0.48
13	propan-2-ol	0.33	0.38	0.56	0.48
14	trifluoroetanol	0.57	0.62	0.25	0.48
15	phenol	0.60	0.58	0.30	0.25
16	*m*-cresol	0.57	0.59	0.31	0.26
17	*p*-cresol	0.57	0.58	0.34	0.32
18	diethyl ether	0.00	0.00	0.45	0.31
19	furan	0.00	0.00	0.13	0.53
20	methylamine	0.16	0.20	0.58	0.25
21	propylamine	0.16	0.22	0.61	0.55
22	butylamine	0.16	0.22	0.61	0.60
23	dimethylamine	0.08	0.11	0.66	0.59
24	diethylamine	0.08	0.09	0.69	0.72
25	aniline	0.26	0.29	0.41	0.28
26	acetamide	0.54	0.55	0.68	0.65
27	butilamide	0.56	0.52	0.61	0.65
28	N-methylacetamide	0.40	0.29	0.72	0.66
29	N,N-methylacetamide	0.00	0.00	0.78	0.78
30	acetaldehyde	0.00	0.00	0.45	0.46
31	benzaldehyde	0.00	0.00	0.39	0.44
32	butanone	0.00	0.00	0.51	0.46
33	pentan-2-one	0.00	0.00	0.51	0.46
34	acetophenone	0.00	0.00	0.48	0.41
35	benzophenone.	0.00	0.00	0.50	0.45
36	methyl acetate	0.00	0.00	0.45	0.47
37	ethyl acetate	0.00	0.00	0.45	0.48
38	methyl benzoate	0.00	0.00	0.46	0.42
39	ethyl benzoate	0.00	0.00	0.46	0.46
40	benzonitrile	0.00	0.00	0.33	0.30
41	acetonitrile	0.07	0.00	0.32	0.35
42	propene	0.00	0.00	0.07	0.07
43	toluene	0.00	0.00	0.14	0.17
44	chlorobenzene	0.00	0.00	0.07	0.03
45	butadyene	0.00	0.00	0.10	0.06
46	cyclopentene	0.00	0.00	0.10	0.09
47	cyclopentane	0.00	0.00	0.00	0.02
48	cycloheptane	0.00	0.00	0.00	0.04
49	chloroethane	0.00	0.00	0.10	0.08
50	trichloromethane	0.00	0.00	0.00	0.05
51	nitromethane	0.06	0.04	0.31	0.40
52	nitroethane	0.02	0.04	0.33	0.38
53	ethylthiol	0.00	0.00	0.24	0.28
54	n−propylthiol	0.00	0.00	0.24	0.20
55	diethyl sulfide	0.00	0.00	0.29	0.28

variety of chemical situations. The considered compounds are listed in Table 2 with their identification number and with the experimental and theoretical α and β values that will be discussed in the following.

The statistical significance of the theoretical descriptors was determined by the root-mean-square, rms, and by the correlation coefficient, R^2.

RESULTS

A first preliminary analysis of the effective atomic charges for some representative molecules of the sample is reported

ABRAHAM HYDROGEN BOND ACIDITY AND BASICITY FACTORS

J. Chem. Inf. Model., Vol. 45, No. 2, 2005 **331**

Table 3. Comparison of Group Charges (Atomic Units) for Three Molecules[a]

group/atom	Mulliken	Löwdin	this paper
Acetic Acid			
CH₃	0.10	0.12	0.027
C	0.79	0.34	0.013
=O	−0.69	−0.45	−0.056
-O-	−0.73	−0.43	−0.024
H	0.53	0.42	0.041
Acetaldehyde			
CH₃	0.07	0.09	0.027
CH	0.55	0.32	0.035
=O	−0.62	−0.41	−0.061
Methylamine			
CH₃	0.22	0.11	0.025
H	0.34	0.27	0.014
N	−0.90	−0.65	−0.052

[a] The Mulliken and Löwdin charges arises from the electronic density function obtained by SCF−HF−PCM/6-31G* calculation. Our charges are obtained by the charge distribution on the cavity surface by eq 7.
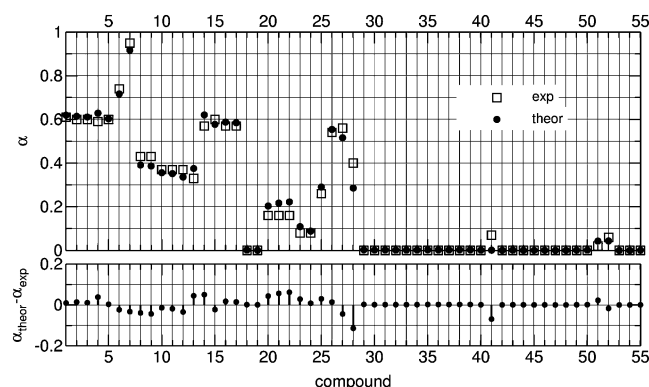
Figure 1. Computed and experimental α values for the considered sample.

in Table 3 through a comparison with the common Mulliken and Löwdin atomic charges obtained by the one-electron density function yielded by the PCM calculation. Our charges are apparently one order of magnitude smaller than the other ones. Nevertheless they preserve the same sign for all groups and, most important for the successive use of the atomic effective charges, the most negative groups (oxygen and nitrogen) are the same in all cases. Some systematic differences are concerned with the methyl charges which are found to be relatively higher with respect to the charges on the remaining molecular groups. For instance in methylamine the net negative charge on nitrogen is neutralized by 40% by CH₃, whereas for Mulliken and Löwdin charges this value is around 20%. For acetic acid our charges agree with those of Löwdin in predicting the sp² oxygen as the most negative atom and, possibly, that involved in the HB. This satisfactory behavior is in agreement with experimental evidences and reinforces the reliability of our effective charges, to be used as basic quantities for the calculation of HB molecular properties.

Figure 1 contains a graphical representation of the comparison between theoretical and experimental α values; the same data are also reported in Table 2. The two sets of data are reported together with their difference, to appreciate the qualities and the defects of the present methodology. For an easier analysis, the compounds are grouped in families.

**Table 4.** Atomic Factors by Eq 8

acidic factors		basicity factors	
H (bonded to O)	15.290	O (sp³)	7.595
H (bonded to S)	0.000	O (sp²)	7.535
H (bonded to N)	7.501	C	3.309
C	0.024	N (sp³)	10.587
		N (sp²)	9.119
		S	9.497
		F	3.036
		Cl	1.810
		Br	0.500

The statistical analysis gives an rms of 0.032 and a R^2 of 0.992. By leaving out the compounds with null HB acidity, which of course tend to improve the quality of the global statistic descriptors, these values change to 0.038 and 0.988, respectively. The quality of the rms is to be judged in the light of the experimental standard deviation reported in the original paper of 0.09. However, as suggested by Platts[11] who considered the successive developments in this field, an error within 0.06 units can be considered adequate for the use of the theoretical α values as reliable descriptors in LFER's. With the exception of *N*-methylacetamide, which shows a too low HB acidity, all of the differences with experiment are within 0.08 units.

The resulting atomic factors which minimize the standard deviation (eq 8) are reported in Table 4. As concerns those relevant for α, it is apparent that the hydrogen atomic factor is greater for H−O bonds than for H−N bonds, whereas for H−S bonds it is very small. This is in agreement with the chemical insight that the electronegativity of the atom bonded to hydrogen is decisive for the acidic strength. The dependence of the hydrogen factors to the bonded atom seems to indicate that the significant positive charge on the hydrogen atom induced by the electronegative bonded atom is not sufficient to reproduce the correct acidity strength. This confirms that, as extensively discussed above, the neutral molecule AH (eq 1) cannot contain all the needed information related to α. Some of them are contained in the A−H··B complex where the bonded atom bears a net charge, greater than that in the reactant. Thus the atomic factors have to be considered as empirical corrections related to the relative stability of the complex.

A similar conclusion was reached by Platts[10] who tried to correlate the charge on the acidic hydrogen atoms with the α value, even in conjunction with the charge on the bonded atom. The best correlation coefficient for the Atoms in Molecules population analysis was 0.562, a rather low value indicating a very low correlation between these two quantities.

To consider a large variety of molecules we have included also cyanides and nitro compounds which have the peculiarity of showing small acidity, despite the absence of hydrogen atoms bonded to a heteroatom. Their HB acidity arises from hydrogen atoms in the methyl group, which in the UA approach in enclosed in a single spherical cavity centered on the carbon atom and, consequently, is inadequate to reveal specific properties they are related to. These hydrogen atoms cannot take any atomic effective charge and are thus treated in a different way with respect to the acidic hydrogen atoms bonded to a heteroatom. Therefore in these cases is the positive charge on the CH₃ group and the acidity factor of the carbon which account for the acidic capacity of the
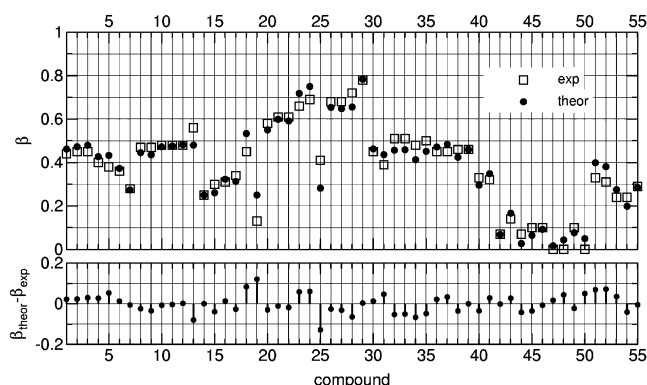
Figure 2. Computed and experimental β values for the considered sample.

hydrogen atoms in the methyl and methylene groups. This is of course a drastic simplification, but in view of the low experimental α values for these compounds, it can be considered an acceptable approximation. Despite the small value of the acidic atomic factor for carbon (about 0.02), the nitro compounds 51 and 52 are seen to possess some acidic capacity in accord with experiment, whereas no acidity is found for cyanides (compound 41).

Despite the good statistical error of the whole theoretical α values, it is somewhat unpleasant that the *N*-methylacetamide acidic capacity is underestimated by 0.1 units. On the other hand its α values are about one-half that of acetamide and, considering the relatively small effect expected for the methyl group bonded to nitrogen, it appears to be reasonable and in line with the chemical statistics. The disagreement arises from the fact that its experimental α is about 4/5 that of acetamide and this is somewhat surprising. This disagreement should be partially corrected by restoring the attenuation factor for the two acidic hydrogens; but this seemed a correction in contrast with chemical reasonableness, and it was rejected.

In the case of carboxylic acids, alcohols and amines the agreement is remarkable. Notice that for the first two classes of compounds the atomic charges are mainly translated into HB acidity by the (single) atomic factor of the hydrogen atoms bonded to oxygen. For both classes the empirical correction for the unbalanced inductive effect of the halogen atoms leads to accurate results.

The corresponding data for the HB basicity factors are reported in Figure 2 and in Table 2. In this case the rms is 0.046 and the R^2 is 0.972.

The theoretical results of the carboxylic acids reproduce correctly the decreasing HB basicity on going from aliphatic to aromatic and halogenated compounds. The same holds for the alcohols with the exception of the secondary one, whose theoretical β is the same of the primary ones in contrast with the experimental increase of the HB basicity. The accuracy is remarkable for aldehydes, ketones, and several other classes of compounds including halogenated hydrocarbons, nitro and sulfur compounds. For the amines the increase of HB basicity on going from the primary to the secondary case is accurately reproduced by the theory, whereas a marked disagreement is found for the aromatic amine whose β value is underestimated by about 0.1 units. Since the HB basicity of the aliphatic amines is correctly reproduced, it seems that the disagreement could be found

in an excessive effect of the aromatic ring. Despite the nonplanar conformation of aniline, the aromatic ring should stabilize the free molecule through mesomeric effect which delocalizes the lone pair electrons on the ring. Since smaller effects are expected for the hydrogen bonded system, where conjugation is not effective for quaternary nitrogen atoms, the inclusion of physically acceptable empirical corrections should lead to a further decrease of the HB basicity. The marked difference between aliphatic and aromatic amines finds however a correspondence in the basicity of amines in water where the behavior is much more different: pK_b is 3.4 for methylamine and 9.4 for aniline. The other most relevant difference between experiment and theory is found for the two ethers included in the molecular set, whose theoretical β's result overestimated. It is however remarkable that the strong decrease going from diethyl ether to furan is correctly reproduced by the theory.

It is seen from Table 3 that the atomic factors for oxygen are completely independent from the hybridization and for nitrogen the dependence is small. The consequence is that the minimum value of the functional (eq 6) changes little if the dependence from hybridization is removed and only one atomic factor is employed for oxygen and nitrogen. Indeed in this case the rms does not change appreciably, and the correlation coefficient decreases from 0.972 to 0.971. This finding allows us to decrease the number of fitting parameters and to make our method more robust from a chemical point of view. The finding that the AF's of nitrogen atom are higher than those of oxygen atom is in line with the common chemical insight that nitrogen is more basic than oxygen.

COMPARISON WITH PREVIOUS CALCULATIONS

To our knowledge, Politzer and Murray[34,35] were the first who looked for correlation between the experimental Abraham descriptors and ab-initio molecular properties. The minimum value of the electrostatic potential and the maximum value at an isodensity surface around the molecule were found to correlate with the experimental β and α values, respectively. Significant correlations were found along several classes of compounds, although the different fitting parameters precluded the possibility of obtaining a general relation.

More extended works based on DFT calculations were subsequently performed by Platts and co-workers[10,12] employing a much larger number of molecular properties other than the electrostatic potential. They found that, whereas single molecule calculations were able to correlate with the experimental α values, good correlations could be obtained for β by considering the molecule-HF complex. The latter technique has some resemblance with the present one, since both approaches make use of a 'probe' to extract physical information from the molecular response. However, the obtained correlation coefficients were smaller than those obtained here.

The peculiarity of the present approach is the use of a full QM molecular wave function to gain effective atomic charges, which can be connected with the experimental acidity and basicity factors, through a number of atomic factors embodying the ability of the different atoms to form hydrogen bond. Nevertheless, the atomic charges computed here have no rigorous physical meaning, and this is the main

ABRAHAM HYDROGEN BOND ACIDITY AND BASICITY FACTORS

J. Chem. Inf. Model., Vol. 45, No. 2, 2005 **333**

difference with respect the two approaches discussed above. In other words, the price we have to pay to obtain reliable results for a large number of molecular classes is the giving up of using rigorous QM molecular observables as the key quantities to be correlated with the experimental α and β values.

CONCLUSIONS

In this paper we have presented a new technique for the evaluation of two important molecular descriptors widely used in empirical LFER's. The method starts with a quantum mechanical Hartree−Fock calculation with moderate basis set, including a Reaction Field method to probe the response properties of the molecule under examination. The effective charges obtained from the density charge on the cavity surface allow rather accurate estimate of the HB acidity and basicity factors through the inclusion of semiempirical parameters for each relevant atom or group of atoms. These parameters are called atomic factors and are thought of as the elements which supply information about the product of the reaction (eq 1). The method is cheap and can be easily extended to large molecules with a moderate increase of computational effort.

The quality of the results for 55 compounds seems to be rather good in order to allow these values to be included in LFER with good confidence. This confirms that reaction field methods can be successfully used to predict response properties of molecular systems as regards their ability to form HB complexes. The extension of such a method of calculation to other molecular descriptors is in progress in our laboratory.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Raevsky, O. A.; Schaper, K. J. Quantitative estimation of hydrogen bond contribution to permeability and absorption processes of some chemicals and drugs. *Eur. J. Med. Chem.* **1998**, *33*, 799−807.

(2) Spange, S.; Reuter, A. Hydrogen-Bond Donating and Dipolarity/ Polarizability Properties of Chemically Functionalized Silica Particles *Langmuir* **1999**, *15*, 141−150.

(3) Vitha, M. F.; Carr, P. W. Study of the Polarity and Hydrogen-Bond Ability of Dodecyltrimethylammonium Bromide Micelles by the Kamlet−Taft Solvatochromic Comparison Method. *J. Phys. Chem. B* **1998**, *102*(11), 1888−1895.

(4) Welke, B.; Ettlinger, K.; Riederer, M. Sorption of Volatile Organic Chemicals in Plant Surfaces. *Environ. Sci. Technol.* **1998**, *32*, 1099−1104.

(5) Berthelot, M.; Laurence, C.; Foucher, D.; Taft, Partition R. W. Coefficients and intramolecular hydrogen bonding. 1. The hydrogen-bond basicity of intramolecular hydrogen-bonded heteroatoms. *J. Phys. Org. Chem.* **1996**, *9*(5), 255−261.

(6) Taft, R. W.; Gurka, D.; Joris, L.; Schleyer, P. von R.; Rakshys, J. W. Studies of hydrogen-bonded complex formation with p-fluorophenol. V. Linear free energy relationships with OH reference acids. *J. Am. Chem. Soc.* **1969**, *91*, 4801−4808; Mitskym J.; Joris L. Taft R. W. Effects of polar aprotic solvents on linear free-energy relations in hydrogen-bonded complex formation. *J. Am. Chem. Soc.* **1972**, *94*, 3438−3442.

(7) Raevsky, O. A. In *QSAR in Drug Design and Toxicology'*, ed D. Hadzi and B. Jerman-Balzic, Elsevier: Amsterdam, 1987

(8) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Morris, J. J.; Taylor, P. J. Hydrogen-bonding. 10. A scale of solute hydrogen-bond passivity using log *K* values for complexation in tetrachloromethane. *J. Chem. Soc., Perkin Trans.* **1990**, *2*(4), 521−529.

(9) Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **1993**, *22*(2), 73−83.

(10) Platts, J. A. Theoretical prediction of hydrogen bond donor capacity. *Phys. Chem. Chem. Phys.* **2000**, *2*(5), 973−980

(11) Dearden, J. C.; Ghafourian, T. Hydrogen Bonding Parameters for QSAR: Comparison of indicator variables, hydrogen bond counts, molecular orbital and other parameters. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 231−235

(12) Lamarche, O.; Platts, J. A. Theoretical prediction of the hydrogen-bond basicity pK_{HB}. *Chem. Eur. J.* **2002**, *8*(2), 457−466.

(13) Platts, J. A. Theoretical prediction of hydrogen bond basicity. *Phys. Chem. Chem. Phys.* **2000**, *2*(14), 3115−3120.

(14) Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. Estimation of molecular linear free energy relation descriptors using a group contribution approach. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 835−845

(15) Zissimos A. M.; Abraham M. H.; Klamt A.; Eckert F.; Wood J. A Comparison between the two general sets of linear free energy descriptors of Abraham and Klamt. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1320−1331.

(16) Klamt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of salvation phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(17) Famini, G. R.; Wilson, L. Y. Using theoretical descriptors in linear free energy relationships: characterizing several polarity, acid and basicity scales. *J. Phys. Org. Chem.* **1999**, *12*, 645−653.

(18) Lamarche, O.; Platts, J. A.; Hersey, A. Theoretical prediction of partition coefficients via molecular electrostatic and electronic properties. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 848−855.

(19) Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **1936**, *58*, 1486−1493.

(20) Warshel, A. Calculation of chemical processes in solutions. *J. Phys. Chem.* **1979**, *83*, 1640−1652.

(21) Tomasi, J.; Persico, M. Molecular interaction in solutions: an overview of methods based on continuous distribution of the solvent. *Chem. Rev.* **1994**, *94*, 2027−2094.

(22) Miertus, S.; Scrocco, E.; Tomasi J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117−129.

(23) Barone, V.; Cossi, M.; Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **1997**, *107*, 3210−3221.

(24) Cacelli, I.; Ferretti, A. Solvent effect on the optical properties of [(NH3)(5)Ru-pyrazine](+m) (*m*=2, 3) complexes by ab initio calculations. *J. Chem. Phys.* **1998**, *109*, 8583−8590.

(25) Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. First principles calculations of aqueous p*K*(a) values for organic and inorganic acids using COSMO−RS reveal an inconsistency in the slope of the p*K*(a) scale. *J. Phys. Chem. A* **2003**, *107*, 9380−9386.

(26) Mulliken, R. S. Electronic population analysis on LCAO-MO molecular wave functions. *J. Chem. Phys.* **1955**, *23*, 1833.

(27) Löwdin, P. O. Correlation problem in many-electron quantum mechanics. I. Review of different approaches and discussion of some current ideas. *Adv. Chem. Phys.* **1970**, *5*, 185−199.

(28) Mayer, I. Charge, bond order and valence in the AB initio SCF theory. *Chem. Phys. Lett.* **1983**, *97*, 270.

(29) Reed, A. E.; Curtiss, L. A.; Weinhold F. Intermolecular interactions from a natural orbital, donor−acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.

(30) Bader, R. F. *Atoms in Molecules − A Quantum Theory*, Oxford University Press: Oxford, 1990.

(31) Oliferenko, A. A.; Oliferenko, P. V.; Huddleston, J. G.; Rogers, R. D.; Palyulin, V. A.; Zefirov, N. S.; Katritzky, A. R. Theoretical Scales of Hydrogen Bond Acidity and Basicity for Application in QSAR/ QSPR Studies and Drug Design. Partitioning of Aliphatic Compounds. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1042−1055.

(32) Frisch, M. J.; et al. *Gaussian 98*, Rev. A.6, Gaussian, Inc., Pittsburgh, PA (USA), 1998.

(33) Schmidt, M. W. et al. *Gamess*. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(34) Murray, J. S.; Politzer, P. Correlation between the solvent hydrogen bond acceptor parameter β and the calculated molecular electrostatic potential. *J. Ort. Chem.* **1991**, *56*, 3734−3737.

(35) Murray, J. S.; Politzer, P. Correlation between the solvent hydrogen bond donating parameter α and the calculated molecular surface electrostatic potential. *J. Ort. Chem.* **1991**, *56*, 6715−6717.