Prediction of the Intrinsic Hydrogen Bond Acceptor Strength of Organic Compounds by Local Molecular Parameters

Johannes Schwöbel,[†,‡] Ralf-Uwe Ebert,[†] Ralph Kühne,[†] and Gerrit Schüürmann*[,†,‡]

UFZ Department of Ecological Chemistry, Helmholtz Centre for Environmental Research, Permoserstrasse 15, 04318 Leipzig, Germany, and Institute for Organic Chemistry, Technical University Bergakademie Freiberg, Leipziger Strasse 29, 09596 Freiberg, Germany

Received January 30, 2009

A quantum chemical model has been developed for predicting the hydrogen bond (HB) acceptor strength of monofunctional organic compounds from electronic ground-state properties of the single molecules. Local molecular parameters are used to quantify electrostatic, polarizability, and charge transfer components to hydrogen bonding, employing the ab initio and density functional theory levels HF/6−31G** and B3LYP/6−31G**. The model can handle lone pairs of intermediate and strong HB acceptor heteroatoms (N, O, S) as well as of weak HB acceptor halogens (F, Cl, Br) and includes also olefinic, alkyne, and aromatic π-bonds as weak HB acceptor sites. The model calibration with 403 compounds and experimental values for the Abraham HB acceptor strength B yielded squared correlation coefficients r^2 around 0.95, outperforming existing fragment-based schemes. Model validation was performed applying a leave-50%-out procedure, yielding predictive squared correlation coefficients q^2 of around 0.95 for the subsets that both cover the whole chemical domain as well as (almost) the whole target value range of the data set.

INTRODUCTION

The quantification of the hydrogen bond[1,2] strength is of considerable practical relevance in chemical, biochemical, and pharmaceutical issues.[3,4] Hydrogen bonding has been exposed to be one of the key parameters in the partitioning and allocation of bioactive compounds in the environment.[5,6]

Hydrogen bonding is a directional interaction between a hydrogen bond donor and a hydrogen bond acceptor site, with strengths between van der Waals and covalent interactions. The properties of the acceptor site mainly contribute to the geometrical and chemical behavior, especially in the case of strong hydrogen bonds.[7,8]

Taft proposed first an experimental pK_{HB} hydrogen bond (HB) basicity scale, related to the Gibbs free energy ΔG^0 of complex formation of HB acceptors with the HB donor 4-fluorophenol in CCl_4 at 298 K.[9] Abraham and co-workers developed a more general B scale.[5] The scale is defined such that zero expresses no hydrogen bonding ability, e.g. pure alkanes, and one refers to the strong base hexamethylphosphorous triamide.

So far, prediction of the HB strength B from molecular structure is confined to methods employing only topological information. The Platts model[10] and its update available only as commercial Absolv ADME model[11] are based on fragments and have been subject to comparative analyses of 2D models for predicting physicochemical properties.[12,13]

Existing approaches from 3D structure follow the supermolecular approach, evaluating the formation energies of HB donor−acceptor complexes.[14] Platts and co-workers used both hydrofluoric acid[15,16] and hydrofluoric acid dimers[17] to model HB basicity by DFT methods with reasonable accuracy. They correlated basicity scales and HB binding energy or Gibbs free energy among other parameters, considering geometry and charge density changes. Raub, Marian, and co-workers used an empirical scoring function to calculate H-bond interaction energies based on *ab initio* derived increments.[3,18]

The goal of the present study is to develop a quantum chemical method for predicting the site-specific HB acceptor strength B from the 3D structure of the isolated compound. This approach reduces the computational expense, and problems associated with the basis set superposition errors (BSSE)[19] are avoided from the beginning. For the electronic characterization of the HB acceptor sites, local molecular parameters, originally developed by Klamt,[20,21] together with charge descriptors are employed. Recently, a new model based on similar parameters demonstrated good performance for predicting the H-bond donor strength.[22] As demonstrated below, our local-parameter approach yields also good statistics for predicting the HB acceptor strength B.

MATERIAL AND METHODS

Data Set. From various literature sources (see the Supporting Information), experimental values for the Abraham acceptor strength parameter B have been collected for 403 aliphatic and aromatic molecules that all contain only one HB acceptor site. The compound set covers various compound classes and contains oxygen, nitrogen, sulfur, and halogen atoms with appropriate lone pairs and weak carbon π-bonds as HB acceptor sites.

Computational Details. The geometries of all molecules have been optimized at the *ab initio* HF/6−31G** level and

* Corresponding author phone: +49-341-235-1262; fax: +49-341-235-1785; e-mail: gerrit.schuurmann@ufz.de.
† UFZ Helmholtz Centre for Environmental Research.
‡ Technical University Bergakademie Freiberg.

10.1021/ci900040z CCC: $40.75
Published on Web 03/19/2009

the density functional theory B3LYP/6−31G** level,[23,24] respectively, employing Gaussian 03.[25] For all compounds, frequency analysis was carried out to confirm that correct ground-state geometries had been found. Charge analysis and Natural Bond Orbital (NBO) analysis methods were applied by means of the NBO 3.1 implementation[26] in Gaussian 03. Electrostatic potential derived charges (ESP) evaluated at a large number of grid points[27,28] were calculated using the Merz−Singh−Kollmann scheme.[29] For the calculation of effective donor and acceptor energies as specified in the next section, we used an in-house Fortran 77 program, which searches for possible HB acceptor sites and quantifies their intrinsic acceptor strengths. Nonlinear parameter calibration was performed by the Levenberg−Marquardt algorithm, where the Jacobian is then calculated by a forward-difference approximation. This algorithm is included in the freely available Fortran MINPACK library.[30]

Application Domain. The 403 training set compounds cover the following HB acceptor site atom types: olefinic, alkyne and aromatic C, halogen (F, Cl, Br) attached to aliphatic or aromatic C, hydroxyl O attached to aliphatic and aromatic C, ether O, carbonyl O (aldehydes and ketones), heteraromatic O (furans), aliphatic and aromatic amine N, heteroaromatic N (pyridines, pyrroles), thiol S attached to aliphatic or aromatic C, thioether S, and heteroaromatic S (thiophenes).

Besides unsaturated hydrocarbons, only compounds with one heteroatom as the HB acceptor site were used for the model development and calibration. Accordingly, the possible impact of intramolecular hydrogen bonding on the HB acceptor strength B is currently not accounted for. Moreover, molecular structure representation was confined to the minimum-energy geometries, thus neglecting a possible dependence of calculated B on conformation.

In the data set, B ranges from 0.05 to 0.79. Accordingly, the target value range of the model is from essentially 0 to about 0.8, and prediction results for compounds of the chemical domain significantly larger than 0.85 would be doubtful.

Model Validation. The prediction capability was evaluated following the previously introduced procedure of simulated external validation[31] (extended leave-50%-out), augmented by a stratification according to HB acceptor atom types. To this end, the data set was divided into two groups of (almost) the same size, group I and group II. The compounds were ordered by the HB acceptor atom type, followed by their experimental Abraham descriptor B values. Then, the 202 compounds with odd numbers were allocated to group I and the 201 compounds with even numbers to group II. Subsequently, for each group nonlinear calibration of the HB acceptor model was performed as described above. Finally, the model trained with group I was used to predict all B values of group II, and correspondingly all group I B values were predicted by the group II model. In this way, both the robustness (stability of calibration statistics and regression coefficients) and prediction capability (q^2 statistics,[32] see below) can be assessed, using two (temporary) prediction sets that both cover the whole target value range and chemical domain of the data set and at the same time are nontrivial in their size.

Statistical Evaluation. The calibration performance (goodness of fit) of the models to predict B from calculated molecular parameters was assessed through calculation of the squared correlation coefficient

$$r^2 = 1 - \frac{\sum_i^n (y_i^{\text{fit}} - y_i^{\text{obs}})^2}{\sum_i^n (y_i^{\text{obs}} - y^{\text{mean}})^2} \quad (1)$$

and the associated root-mean-square error (*rms*). Here, y_i^{fit}, y_i^{obs}, and y^{mean} denote the regression-fitted and observed target values (in our case: B) and the observed mean, and n the number of respective data. For the prediction performance of the group I and group II models, the predictive squared correlation coefficient[32]

$$q^2 = 1 - \frac{\sum_i^n (y_i^{\text{pred}} - y_i^{\text{obs}})^2}{\sum_i^n (y_i^{\text{obs}} - y^{\text{mean}})^2} \quad (2)$$

was used. In eq 2, y_i^{pred} denotes the predicted (not fitted) target value, and y^{mean} is the mean of observed B values of the test set (group I or group II) used for evaluating the prediction performance.[32]

DEVELOPMENT OF THE HYDROGEN BOND ACCEPTOR STRENGTH MODEL

Hydrogen Bond Interaction Terms. The model considers the following components of hydrogen bonding: electrostatic interaction (ES), polarizability (PL), and charge transfer (CT) that represents a covalent contribution. Starting from the recently developed model to predict the Abraham parameter A (HB donor strength) from molecular structure,[22] a corresponding model equation was set up for the HB acceptor strength B:

$$B(s) = c_{\text{ES}} \cdot Q(s) + c_{\text{PL}} \cdot \eta(s) + c_{\text{CT}} \cdot EE_{\text{occ}}(s) + c \quad (3)$$

In eq 3, s denotes the HB acceptor site of the compound, and c_{ES}, c_{PL}, and c_{CT} are regression coefficients of the (quantum chemically calculated) ES, PL, and CT contributions to B in the Abraham scale, and c is the associated intercept.

Calculation of the ES component proceeds through evaluation of the net atomic charge $Q(s)$ at the HB acceptor site s. The underlying assumption is that the dominating part of the electrostatic interaction is localized at this electron-rich site that provides a Coulomb attraction of the (not explicitly considered) donor hydrogen of the HB partner compound. Partial charge is not a quantum chemical observable, so several charge descriptors with different scopes have already been under discussion.[15,33] The wave function based Mulliken population analysis has known deficiencies and varies substantially with the basis set and Hamiltonian.[34] The more refined natural population analysis (NPA) is sensitive to structural changes but more robust with respect to the level of calculation. Both NPA and ESP charges yield physically reasonable results and thus are considered useful in the present context of modeling the ES component of HB acceptor sites.

In eq 3, $\eta(s)$ denotes the local hardness at the HB acceptor site s. It is used to quantify the PL contribution to B, keeping in mind that increasing hardness corresponds to decreasing polarizability. To characterize the site-specific molecular readiness for donating or accepting electronic charge, local hardness is defined as follows:[22]

$$\eta(s) = -\frac{1}{2}(EE_{\mathrm{occ}}(s) - EE_{\mathrm{vac}}(s)) \tag{4}$$

In eq 4, $EE_{\mathrm{occ}}(s)$ denotes the effective donor energy at site s and is constructed through a sum of occupied molecular orbital (MO) energies E_{i}, weighted by exponential terms $w_i(E_{\mathrm{ref}}, s)$ that contain a reference energy E_{ref} associated with H bond formation:

$$EE_{\mathrm{occ}}(E_{\mathrm{ref}}, s) = \frac{\sum_{i=\mathrm{HOMO}}^{1} E_i \cdot w_i(E_{\mathrm{ref}}, s)}{\sum_{i=\mathrm{HOMO}}^{1} w_i(E_{\mathrm{ref}}, s)} \text{ with}$$

$$w_i(E_{\mathrm{ref}}, s) = p_i(s) \cdot \exp\left(-\frac{E_i}{E_{\mathrm{ref}}}\right) \tag{5}$$

Correspondingly, the effective acceptor energy, $EE_{\mathrm{vac}}(s)$, involves unoccupied MO energies E_{k} and a reference energy E_{ref} suitable to scale E_{k} properly:

$$EE_{\mathrm{vac}}(E_{\mathrm{ref}}, s) = \frac{\sum_{k=LUMO}^{\max} E_k \cdot w_k(E_{\mathrm{ref}}, s)}{\sum_{k=LUMO}^{\max} w_k(E_{\mathrm{ref}}, s)} \text{ with}$$

$$w_k(E_{\mathrm{ref}}, s) = p_k(s) \cdot \exp\left(-\frac{E_k}{E_{\mathrm{ref}}}\right) \tag{6}$$

In eqs 5 and 6, the population p_i (p_k) of occupied (unoccupied) molecular orbital i (k) is quantified by taking twice the sum of the square of the LCAO-MO (linear combination of atomic orbital - molecular orbital) coefficients $c_{\mu i}$ ($c_{\rho k}$) at an atomic center r:

$$p_i(r) = 2 \cdot \sum_{\mu(r)} (c_{\mu i})^2;\ p_k(r) = 2 \cdot \sum_{\rho(r)} (c_{\rho k})^2 \tag{7}$$

In these and the following equations, Greek indices μ, ρ... denote the atomic orbitals, while Latin letters i,k, ... run over the molecular orbitals of an atomic center r.

If an acceptor site s contains not only one but n_{A} atomic centers r, the associated population is

$$p_i(s) = 2 \cdot \frac{\sum_r \sum_{\mu(r)} (c_{\mu i})^2}{n_{\mathrm{A}}};\ p_k(s) = 2 \cdot \frac{\sum_r \sum_{\rho(r)} (c_{\rho k})^2}{n_{\mathrm{A}}} \tag{8}$$

For example, the π-electrons of toluene as weak HB acceptor system are delocalized over 6 carbon atoms, thus yielding $n_{\mathrm{A}} = 6$. This equation works, if each acceptor-site atom is described by the same number of atomic orbitals; otherwise, a normalization through weighting by the different number of atomic orbitals has to be introduced. Because eqs 7 and 8 ignore the off-diagonal contributions to the electron density (which in turn cannot be uniquely allocated to individual atomic sites, and which would be zero in semiempirical models that neglect diatomic differential overlap), $p_{\mathrm{i}}(s)$ and $p_{\mathrm{k}}(s)$ provide only approximate characterizations of the electron density and electron space available at atomic site $r(s)$ when employing *ab initio* and DFT levels of calculation.

The third component of eq 3 represents the CT contribution to the HB acceptor strength B. Here, the effective donor energy $EE_{\mathrm{occ}}(s)$ at HB acceptor site s (see eq 5) is used to characterize its readiness for donating electron charge upon hydrogen bonding. The covalent HB component is thus modeled according to the concept of orbital interaction, quantifying the electron donation energy of the HB acceptor site through the respectively calibrated parameter EE_{occ}. We also tried to introduce an additional term with the lone pair or π-bond energy levels, which can be obtained by NBO analysis. However, the model improvement was not statistically significant.

Table 1. Reference Energy for Different Quantum Chemical Methods (in eV)[a]

	E_{ref} (HF/6−31G**)		E_{ref} (B3LYP/6−31G**)	
HB acceptor	EE_{occ}	EE_{vac}	EE_{occ}	EE_{vac}
lone pair	34.07	3.51	12.00	2.51
π-bond	4.06	3.00	10.87	2.50

[a] The effective donor and acceptor energies, EE_{occ} and EE_{vac}, and their associated reference energies E_{ref} are defined in eqs 5 and 6, respectively.

RESULTS AND DISCUSSION

Model Calibration. The parameters of the model were calibrated in the following way: Because the effective donor and acceptor energies, EE_{occ} (eq 5) and EE_{vac} (eq 6), contain nonlinear parameters, their optimization was performed using the Levenberg−Marquardt algorithm. The resultant parameters are listed in Table 1. Different parameters are applied for lone pair and π-bond HB acceptor systems. To our surprise, there is no need to distinguish between atomic centers with one lone pair (nitrogen), with two lone pairs (oxygen) and with several lone pairs (halogens).

The associated coefficients c_{ES}, c_{PL}, c_{CT}, and c are shown in Table 2, again calculated by nonlinear calibration of the data set. Separation of the calculations dependent on atom type was necessary, but within one atom type all respective compound classes (as included in the data set) could be taken into account for the calibration.

ES, PL, and CT Contributions to B. The regression coefficient of the electrostatic component, c_{ES}, is negative for both the HF and B3LYP level of calculation (see Table 2). Thus, increasingly negative charge at the HB acceptor site increases the strength of hydrogen bonding, reflecting the correspondingly increasing Coulomb attraction to the positively charged hydrogen of the (not explicitly considered) HB donor site.

With HF, the local hardness coefficient c_{PL} is positive for all heteroatoms (N, O, S, F, Cl, Br), and negative for carbon, keeping in mind that increasing hardness implies decreasing polarizability (PL). In general, c_{PL} decreases in the following way for the different atom types: F $\gg$ *N* > O $\approx$ S > C, and F $\gg$ Cl > Br. The positive sign would be in accord with considering hydrogen bonding as hard−hard interaction[22]

Table 2. Regression Coefficients of the Hydrogen Bond Acceptor Model To Predict Abraham Parameter *B* from Molecular Structure[a]

HB acceptor type	HF/6−31G**				B3LYP/6−31G**			
	c_{ES}	c_{PL}	c_{CT}	c	c_{ES}	c_{PL}	c_{CT}	c
				NPA Net Atomic Charges				
C	−0.205	−0.105	+0.191	+3.416	−0.000	+0.108	+0.000	−0.751
N	−0.432	+0.098	+0.229	+2.750	−0.130	+0.042	+0.217	+2.373
O	−0.972	+0.080	+0.156	+1.475	−0.313	−0.020	+0.140	+1.902
S	−0.000	+0.079	+0.216	+2.717	−0.047	+0.086	+0.066	+0.316
F	−0.385	+0.220	+0.298	+2.713	−0.667	+0.394	+0.240	−0.426
Cl	−0.372	+0.081	+0.225	+2.716	−0.113	+0.021	+0.087	+0.845
Br	−0.375	+0.077	+0.142	+2.742	−0.000	+0.002	+0.090	+1.095
				ESP Net Atomic Charges				
C	−0.000	−0.125	+0.154	+3.255	−0.000	+0.079	+0.015	−0.347
N	−0.171	+0.030	+0.177	+2.949	−0.001	−0.011	+0.201	+2.672
O	−0.497	+0.035	+0.114	+1.730	−0.201	−0.021	+0.132	+1.922
S	−0.000	+0.050	+0.141	+1.870	−0.000	+0.061	+0.222	+1.965
F	−0.745	+0.126	+0.092	−0.128	−0.714	+0.235	+0.130	−0.385
Cl	−0.497	+0.066	+0.052	+0.003	−0.203	+0.060	+0.120	+0.913
Br	−0.493	+0.065	+0.043	+0.002	−0.000	−0.033	+0.065	+1.105

[a] The associated regression equation is $B(s) = c_{ES} \cdot Q(s) + c_{PL} \cdot \eta(s) + c_{CT} \cdot EE_{occ}(s) + c$ (eq 3), where *s* denotes the HB acceptor site, Q = net atomic charge, η = local hardness, and EE_{occ} = effective donor energy (see eqs 4−8). NPA = natural population analysis, ESP = electrostatic potential.

Table 3. Hydrogen Bond Acceptor Model Statistics[a]

method	*n*	r^2	*rms*	*me*	*mne*	*mpe*
HF (NPA)	403	0.96	0.04	0.03	0.15	0.25
HF (ESP)	403	0.95	0.04	0.03	0.14	0.26
B3LYP (NPA)	403	0.95	0.04	0.03	0.22	0.16
B3LYP (ESP)	403	0.94	0.05	0.04	0.25	0.19

[a] The statistical parameters are as follows: *n* = number of chemicals, r^2 = squared correlation coefficient, *rms* = root-mean-square error, *me* = mean error, *mne* = maximum negative error (largest underestimation), *mpe* = maximum positive error (largest overestimation).

Table 4. Root-Mean-Square Errors *rms* for Different Atom Types[a]

type	*n*	HF (NPA)	HF (ESP)	B3LYP (NPA)	B3LYP (ESP)
all	403	0.040	0.043	0.044	0.047
C	66	0.034	0.034	0.033	0.038
N	88	0.048	0.060	0.047	0.050
O	182	0.039	0.042	0.050	0.053
S	20	0.038	0.031	0.057	0.043
F	10	0.048	0.026	0.034	0.031
Cl	18	0.038	0.010	0.009	0.011
Br	19	0.029	0.015	0.009	0.009

[a] For the HF (Hartree−Fock) and B3LYP calculations, the basis set 6−31G** has been used, employing either NPA (natural population analysis) or ESP (electrostatic potential) for quantifying net atomic charges; *n* = number of compounds.

according to the Pearson concept. The negative sign with C means that electronically soft sites such as π-electron systems prefer to become polarized upon hydrogen bonding. Interestingly, the π-electron carbon coefficient is positive for both NPA and ESP at the DFT level. Note that B3LYP includes electron correlation, which might have significant influence on the description of a delocalized π-system, and which is apparently not reflected by the HF level of calculation. On the other side, HF schemes represent the best 1-electron model possible. However, both levels of computation provide the same performance for C (see Table 4), possibly due to fortuitous error compensation.

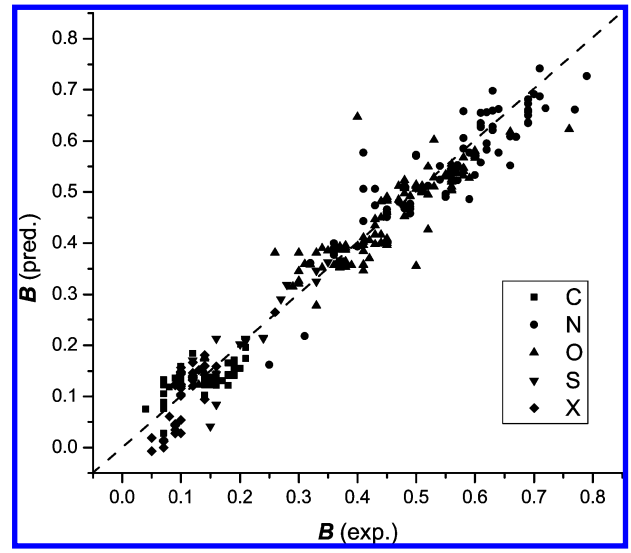

Figure 1. Acceptor model performance (HF/6−31G** + NPA): squares, C; circles, N; triangles up, O; triangles down, S; rhombs, F, Cl, Br.

With DFT, positive c_{PL} values using the NPA net atomic charge scheme can turn into - only small - negative values using the ESP scheme. Obviously there is some interdependence of the net atomic charges and the polarizability descriptor. Indeed, the intercorrelation between the ES term and the PL term is 68% (ES-CT: 29%; PL-CT: 32%) using NPA charges and 53% (ES-CT: 20%; PL-CT: 32%) using ESP charges. So, about two-thirds of the variation of the electrostatic term can be explained by the variation of the polarizability term. Nevertheless, both ES and PL are needed for both physical and statistical reasons to describe the HB strength sufficiently well, and the derived regression coefficients appear to be quite robust as is demonstrated below through leave-50%-out cross validation. A similar situation had been observed for our HB bond donor model.[22]

The coefficient of the charge transfer contribution to *B*, c_{CT}, is positive for each method. It confirms our expectation that with increasing effective donor energy, its interaction with the (not explicitly considered) unoccupied HB donor

Table 5. Model Validation by Data Set Splitting into Group I and Group II[a]

method	group	mode	n	r^2	*rms*	q^2	bias
HF (NPA)	I	training	202	0.955	0.042	n.a.	n.a.
	II	test	201	0.963	0.038	0.962	−0.004
	II	training	201	0.956	0.041	n.a.	n.a.
	I	test	202	0.945	0.046	0.944	0.003
HF (ESP)	I	training	202	0.959	0.040	n.a.	n.a.
	II	test	201	0.953	0.042	0.953	−0.003
	II	training	201	0.948	0.044	n.a.	n.a.
	I	test	202	0.942	0.047	0.941	0.002
B3LYP (NPA)	I	training	202	0.958	0.040	n.a.	n.a.
	II	test	201	0.952	0.043	0.951	−0.004
	II	training	201	0.952	0.043	n.a.	n.a.
	I	test	202	0.944	0.047	0.942	0.004
B3LYP (ESP)	I	training	202	0.956	0.041	n.a.	n.a.
	II	test	201	0.945	0.046	0.944	−0.004
	II	training	201	0.944	0.046	n.a.	n.a.
	I	test	202	0.919	0.056	0.918	0.002

[a] The quantum chemical HF (Hartree−Fock) and B3LYP calculations have been performed with the 6−31G** basis set, generating either NPA (natural population analysis) or ESP (electrostatic potential) net atomic charges. The subsets group I and group II cover both the whole chemical domain and essentially the whole value range of B. The statistical parameters are as follows: r^2 = squared correlation coefficient, *rms* = root-mean-square error, q^2 = predicted squared correlation coefficient. n.a. = not applicable.

Table 6. Chemical Compound Classes and Range of Experimental HB Acceptor Strength[a]

compound class	HB acceptor site	n	B_{min}	B_{mean}	B_{max}
	Weak HB Acceptor Strength				
halogens (aromatically bound)	F, Cl, Br	10	0.05	0.08	0.10
olefins	C (sp^2)	24	0.07	0.09	0.19
halogens (aliphatically bound)	F, Cl, Br	37	0.05	0.11	0.26
alkynes	C (sp)	12	0.04	0.12	0.21
O-heteroaromatics	O (sp^2)	6	0.13	0.14	0.14
S-heteroaromatics	S (sp^2)	2	0.15	0.16	0.16
thiophenols	S (sp^2)	1	0.16	0.16	0.16
aromatic hydrocarbons	C (sp^2)	30	0.14	0.17	0.21
thiols	S (sp^3)	11	0.12	0.21	0.24
	Moderate HB Acceptor Strength				
anisoles	O (sp^3)	7	0.29	0.31	0.34
sulfides	S (sp^3)	6	0.27	0.32	0.37
nitriles	N (sp)	14	0.32	0.36	0.40
phenols	O (sp^2)	42	0.30	0.39	0.50
aldehydes	O (sp^2)	10	0.33	0.44	0.45
	Strong HB Acceptor Strength				
anilines	N (sp^2-sp^3)	20	0.41	0.47	0.59
ethers	O (sp^3)	23	0.26	0.48	0.76
ketones	O (sp^2)	35	0.47	0.52	0.56
alcohols	O (sp^3)	59	0.47	0.53	0.60
N-heteroaromatics	N (sp^2)	22	0.25	0.56	0.70
amines	N (sp^3)	32	0.58	0.66	0.79

[a] B_{min} = minimum experimental B value, B_{mean} = mean of experimental B values, B_{max} = maximum experimental B value.

Table 7. Root-Mean-Square Errors *rms* of the Predicted Abraham Parameter B for Different HB Acceptor Classes[a]

HB acceptor class	B range	n	HF (NPA)	HF (ESP)	B3LYP (NPA)	B3LYP (ESP)
weak	0.00−0.24	132	0.036	0.028	0.030	0.033
moderate	0.25−0.44	84	0.049	0.063	0.051	0.053
strong	0.45−0.60	134	0.035	0.037	0.045	0.049
very strong	>0.60	53	0.047	0.052	0.059	0.058

[a] The underlying quantum chemical models are HF/6−31G** and B3LYP/6−31G**, respectively, employing net atomic charges according to NPA (natural population analysis) or to ESP (electrostatic potential); n = number of compounds.

orbital is increasingly favored, yielding an increasing energy stabilization through hydrogen bonding.

Model Performance. Application of the model employing four method combinations (HF and B3LYP as well as NPA and ESP) to all 403 compounds yields the statistics summarized in Table 3. In Table 4, the root-mean-square errors are listed separately for each atom type. The performance is similar for HF and DFT, with r^2 values between 0.94 (B3LYP-ESP) and 0.96 (HF-NPA). For the latter, the data distribution is shown in Figure 1.

The model performs well for both heteroatom lone pairs and carbon π-bonds. NPA charges appear to be slightly

preferred to ESP charges and are recommended for prediction purposes also because ESP charges are known to have some problems in the context of steric hindrance.

We make some further assumptions for aromatic systems: If a functional group is directly bonded to a conjugated or aromatic system, the heteroatom bears the whole information, although π-bonds serve as weak HB acceptors as well, if they stand alone. If we take a look at heteroaromatic π-systems, only the heteroatom lone pair needs to be inspected.

There are two significant outliers in our model. The first is di-*tert*-butyl ether (experimental value $B = 0.40$, calculated $B = 0.65/0.56$ at the HF/DFT level), where substantial steric hindrance occurs. Keeping in mind that our model is currently confined to quantifying the intrinsic HB acceptor strength, this outlier suggests that future model refinements should address steric hindrance and (in case of multifunctional compounds) intramolecular hydrogen bonding.

The second outlier is the odorous substance eucalyptol (experimental value $B = 0.76$, calculated $B = 0.62/0.54$ at the HF/DFT level). Here, the observed underestimation of B could be partly caused by conformational flexibility, and the ring strain of the molecule might have an additional influence on the HB acceptor strength.

However, both the HF and DFT model outperform Platts' increment method,[10] the latter of which yields the following regression statistics when applied to all 403 compounds: $r^2 = 0.83$, $q^2 = 0.78$, $rms = 0.07$, and $bias = -0.04$.

Comparison of the calculation errors (regression residuals) between HF and DFT yielded an intercorrelation r^2 of only 0.56, showing that both models could be used as components of a consensus model approach to still increase the confidence in predicted values in cases where both models agree.

Validation. As outlined above, the two subsets group I and group II were generated such that each of them covers both (almost) the whole target value range and the whole chemical domain of the data set. The associated calibration and prediction statistics are summarized in Table 5.

For both HF/6−31G** and B3LYP/6−31G** as well as both NPA and ESP net atomic charges, group I and group II yield similar calibration r^2 (always above 0.94) and *rms* values (always lower than 0.05). A corresponding similarity is also observed for the individual regression coefficients (e.g., HF/6−31G** + NPA: $c_{ES}(O) = -1.001$ (group I) vs -0.949 (group II), $c_{PL}(O) = 0.088$ (I) vs 0.074 (II), $c_{CT}(O) = 0.163$ (I) vs 0.150 (II), $c(O) = 1.478$ (I) vs 1.473 (II); other results not shown). It follows that the model appears to be statistically robust, reflecting a sufficiently balanced composition of the data set.

The q^2 results listed in Table 5 (ranging from 0.92 to 0.96) indicate a very good prediction quality for both the group I model (tested with the group II data set) and the group II model (tested with the group I data set). Note further q^2 is close to r^2 for both subsets and does not vary much between group I and group II, again reflecting a good degree of robustness.

Chemical Domain and Target Values Range. In Table 6, the chemical domain and target value range of the model is specified in terms of HB acceptor atom types and associated B values. As can be seen from the table, all unsaturated carbon atoms and halogen atoms and most sulfur groups are weak acceptor groups. As a general trend, functional groups attached to aromatic systems are weaker HB acceptors than the corresponding aliphatic groups: Phenols and anisoles have moderate B values, while alcohols and ethers are compounds with a strong HB acceptor strength. Anilines are strong HB acceptors, but amines are still stronger in this respect. Note further that carbonyl oxygen is a stronger HB acceptor in ketones than in aldehydes. At present, carboxyl groups (carbon acids and amides) and phosphates are not included, because they contain more than one possible acceptor site. However, they will be addressed in a forthcoming publication.

In Table 7, the performance statistics in terms of *rms* errors is specified for the subgroups of weak, moderate, strong, and very strong HB acceptor groups. Here, the classification is not based on the chemical compound class but on the experimental hydrogen bond acceptor strength B. Taking the HF/6−31G** model with NPA net atomic charges as an example, *rms* is below 0.04 for compounds with weak and strong HB acceptor sites and below 0.05 for the ones with intermediate and very strong HB acceptor strength.

Overall, the results demonstrate that the performance of the HB acceptor model is reasonably well balanced across the B target value range. Moreover, *rms* is in the same range (0.05) as the experimental errors are supposed to be, making this approach a promising tool to assess the HB acceptor strength for compounds where respective experimental data are lacking. Possible applications include the comparative evaluation of the HB acceptor strengths of respective site in biomolecules, yielding pertinent information about their HB interaction profile without the need for explicit consideration of HB donor molecules (solvent molecules or e.g. receptor binding groups).

CONCLUSIONS

Decomposition of hydrogen bonding into electrostatic, polarizability, and charge transfer components offers a way to predict its strength from properly calibrated quantum chemical calculations of the latter components. To this end, the use of local parameters enables a site-specific characterization of the hydrogen bond (HB) donor and acceptor strengths. The presently developed model for predicting the HB acceptor strength in terms of the respective Abraham parameter B is based on the same set of parameters as used previously for predicting the HB donor strength, reflecting the common mechanistic basis underlying this approach. The model covers weak, intermediate, and strong HB acceptor sites including π-electron sites consisting of aromatic, olefinic, and alkyne carbon and outperforms existing increment schemes according to calibration statistics. Future refinements may address steric accessibility as well as the influence of intramolecular hydrogen bonding in case of multifunctional compounds.

ACKNOWLEDGMENT

The initial source code for calculating local molecular parameters was provided by Björn Loeprecht, a former member of our group. Financial support was provided by the European Commission through the projects NOMIRACLE (Contract No. 003956) and OSIRIS (Contract No. 037017), which is gratefully acknowledged.

Supporting Information Available: Table listing all 403 compounds together with experimental and predicted values of the Abraham parameter *B*. This material is available free of charge via the Internet at http://pubs.acs.org. A copy of a program to predict the hydrogen bond acceptor strength from information provided through Gaussian output files can be obtained upon request from the corresponding author.

REFERENCES AND NOTES

(1) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, UK, 1997; pp 11−96.

(2) Scheiner, S. *Hydrogen Bonding* A Theoretical Perspective; Oxford University Press: Oxford, UK, 1997; pp 28−37.

(3) Raub, S.; Steffen, A.; Kämper, A.; Marian, C. M. AIScore: Chemically Diverse Empirical Scoring Function Employing Quantum Chemical Binding Energies of Hydrogen-Bonded Complexes. *J. Chem. Inf. Model.* **2008**, *48*, 1492–1510.

(4) Li, I. L.; Mahindroo, N.; Liang, P. H.; Peng, Y. H.; Kuo, C. J.; Tsai, K. C.; Hsieh, H. P.; Chao, Y. S.; Wu, S. Y. Structure-based Drug Design and Structural Biology Study of Novel Nonpeptide Inhibitors of Severe Acute Respiratory Syndrome Coronavirus Main Protease. *J. Med. Chem.* **2006**, *49*, 5154–5161.

(5) Abraham, M. H. Scales of Solute Hydrogen-Bonding: Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, *22*, 73–83.

(6) Abraham, M. H.; Ibrahim, A. Gas to Olive Oil Partition Coefficients: A Linear Free Energy Analysis. *J. Chem. Inf. Model.* **2006**, *46*, 1735–1741.

(7) Laurence, C.; Berthelot, M. Observations on the Strength of Hydrogen Bonding. *Perspect. Drug. Discovery Des.* **2000**, *18*, 39–60.

(8) Scheiner, S. *In Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers, Inc.: New York 1991; pp 165−218.

(9) Taft, R. W.; Abraham, M. H.; Dougherty, R. M.; Kamlet, M. J. The Molecular Properties Governing Solubilities of Organic Nonelectrolytes in Water. *Nature* **1985**, *313*, 384–386.

(10) 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.

(11) *ADME Boxes, version 4.1*; Advanced Pharma Algorithms Inc.: Toronto, ON, Canada, 2008.

(12) Schüürmann, G.; Ebert, R.-U.; Kühne, R. Prediction of Physicochemical Properties of Organic Compounds from 2D Molecular Structure-Fragment Methods vs. LFER Models. *Chimia* **2006**, *60*, 691–698.

(13) Schüürmann, G.; Ebert, R.-U.; Nendza, M.; Dearden, J. C.; Paschke, A.; Kühne, R. Prediction of Fate-Related Compound Properties. In *Risk Assessment of Chemicals. An Introduction, 2nd ed.*; van Leeuwen, K.; Vermeire, T.; Eds.; Springer Science: Dordrecht, Netherlands 2007; pp 375−426.

(14) Thar, J.; Kirchner, B. Hydrogen Bond Detection. *J. Phys. Chem. A.* **2006**, *110*, 4229–4237.

(15) Platts, J. A. Theoretical Prediction of Hydrogen Bond Basicity. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3115–3120.

(16) Lamarche, O.; Platts, J. A. Complementary Nature of Hydrogen Bond Basicity and Acidity Scales from Electrostatic and Atoms in Molecules Properties. *Phys. Chem. Chem. Phys.* **2003**, *5*, 677–684.

(17) Lamarche, O.; Platts, J. A. Theoretical Prediction of the Hydrogen-Bond Basicity pK_{HB}. *Chem. Eur. J.* **2002**, *8*, 457–466.

(18) Raub, S.; Marian, C. M. Quantum Chemical Investigation of Hydrogen-Bond Strengths and Partition into Donor and Acceptor Contributions. *J. Comput. Chem.* **2007**, *28*, 1503–1515.

(19) Simon, S.; Duran, M.; Dannenberg, J. J. How Does Basis Set Superposition Error Change the Potential Surfaces for Hydrogen-Bonded Dimers. *J. Chem. Phys.* **1996**, *105*, 11024–11031.

(20) Klamt, A. Estimation of Gas-Phase Hydroxyl Radical Rate Constants of Organic Compounds from Molecular Orbital Calculations. *Chemosphere* **1993**, *26*, 1273–1289.

(21) Klamt, A. Estimation of Gas-Phase Hydroxyl Radical Rate Constants of Oxygenated Compounds based on Molecular Orbital Calculations. *Chemosphere* **1995**, *32*, 717–726.

(22) Schwöbel, J.; Ebert, R.-U.; Kühne, R.; Schüürmann, G. Modeling the H Bond Donor Strength of -OH, -NH, and -CH Sites by Local Molecular Parameters. J. Comput. Chem. 2009, in press.

(23) Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(24) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pompelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian Inc.: Pittsburgh, PA, 2003.

(26) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO Version 3.1*; Theoretical Chemistry Institute, University of Wisconsin: Madison, 1993.

(27) Ghafourian, T.; Dearden, J. C. The Use of Molecular Electrostatic Potentials as Hydrogen-Bonding-Donor Parameters for QSAR Studies. *Il Farmaco* **2004**, *59*, 473–479.

(28) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.

(29) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431–439.

(30) Moré, J. J.; Garbow, T. K.; Hillstrom, K. E. *User Guide for MINPACK-1*; Argonne National Laboratory Report ANL-80-74; Argonne, IL, 1980.

(31) Boháč, M.; Loeprecht, B.; Damborský, J.; Schüürmann, G. 2002. Impact of Orthogonal Signal Correction (OSC). On The Predictive Ability of CoMFA Models for The Ciliate Toxicity of Nitrobenzenes. *Quant. Struct.-Act. Relat.* **2002**, *21*, 3–11.

(32) Schüürmann, G.; Ebert, R.-U.; Chen, J.; Wang, B.; Kühne, R. External Validation and Prediction Employing the Predictive Squared Correlation Coefficient - Test Set Activity Mean vs. Training Set Activity Mean. *J. Chem. Inf. Model.* **2008**, *48*, 2140–2145.

(33) Hocquet, A.; Toro-Labbé, A.; Chermette, H. Intramolecular Interactions Along the Reaction Path of Keto-Enol Tautomerism: Fukui Functions as Local Softnesses and Charges as Local Hardnesses. *J. Mol. Struct.* **2004**, *686*, 213–218.

(34) Martin, F.; Zipse, H. Charge Distribution in the Water Molecule - A Comparison of Methods. *J. Comput. Chem.* **2005**, *26*, 97–105.

CI900040Z