J Mol Model (2002) 8:95–101
DOI 10.1007/s00894-002-0075-z

ORIGINAL PAPER

Matthias Hennemann · Timothy Clark

A QSPR-Approach to the Estimation of the pK_{HB} of Six-Membered Nitrogen-Heterocycles using Quantum Mechanically Derived Descriptors

Received: 25 October 2001 / Accepted: 7 February 2002 / Published online: 11 April 2002

Abstract Descriptors derived from semiempirical (AM1) molecular orbital calculations have been used to construct a quantitative structure-property relationship (QSPR) for the thermodynamic hydrogen-bond basicity, pK_{HB}, of a series of six-membered aromatic nitrogen-heterocycles. The resulting model uses four-descriptors (the Coulson charge on the nitrogen atom, the energy of the localized nitrogen lone-pair orbital, the *p*-orbital contribution to this MO and an accessibility angle). The model gives r^2_{cv}=0.95 for 51 compounds with a standard deviation between calculation and experiment of 0.13 log units.

Keywords pK_{HB} · Hydrogen bonding · QSPR · AM1

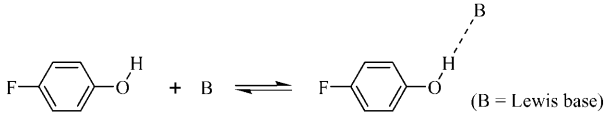
Introduction

Hydrogen bonding is a key element of biologically important intermolecular interactions. Base-pairing in DNA, much of the specificity of enzyme-ligand binding and many molecular recognition phenomena in general can be explained largely on the basis of intermolecular hydrogen bonds. A quantitative description of the hydrogen-bonding ability of a given compound or group within a compound is therefore of fundamental importance for the understanding of biologically important intermolecular interactions and of considerable potential importance for quantitative structure-activity (QSAR-) studies and for scoring functions. Experimentally, a scale of hydrogen-bond basicities, pK_{HB}, has been set up for organic bases, B, (hydrogen-bond acceptors) using 4-fluorophenol as the reference hydrogen-bond donor (Scheme 1). [1] The pK_{HB} is defined as the logarithm of the formation constant of the 1:1 hydrogen-bonded complex with 4-fluorophenol in CCl_4 solvent at 25 °C:

$$K_f = \frac{[4-FC_6H_4OH\cdots B]}{[B]\,[4-FC_6H_4OH]} = \frac{1}{K_{diss}} \quad (1)$$

$$pK_{HB} = -\log_{10} K_{diss} = \log_{10} K_f \quad (2)$$

where K_f and K_{diss} are the formation and dissociation constants, respectively. pK_{HB} Values for 65 six-membered aromatic N-heterocycles were given in the experimental paper [2] and serve as the basis for the current work. Apart from fundamental attempts to understand the exact electronic nature of hydrogen bonding [3], there have been several quantitative structure-property relationship (QSPR) studies designed to predict hydrogen-bonding abilities. Kollman et al. [4, 5] showed that the calculated hydrogen bond energies can be related to the electrostatic potential at fixed distances from hydrogen-bond donors and acceptors. Murray and Politzer [6, 7] pointed out the relationship between surface electrostatic potential properties and later [8] the average local ionization energy to both hydrogen-bonding ability and solute-induced shifts in the methanol O*H*-stretching frequency. Kenny [9] investigated the dependence of the hydrogen-bond basicity on electrostatic potential and field values for use in CoMFA-analyses. Questel et al. [10] used AM1 and PM3 semiempirical calculations to predict the hydrogen bond basicities of nitriles. They concluded that the properties calculated with AM1 correlate significantly better with the experimental values, despite the fact that AM1 severely underestimates the hydrogen-bond enthalpies. Proutiére et al. [11, 12] related experimental dipole moments and frequency shifts to hydrogen-bond basicities and, more recently, Laurence, Questel et al. [13] found a correlation with the minimum electrostatic potential at the molecular surface point



Scheme 1 Equilibrium reaction of 4-fluorophenol and a Lewis base B

M. Hennemann · T. Clark (✉)
Computer-Chemie-Centrum
der Friedrich-Alexander-Universität Erlangen-Nürnberg,
Nägelsbachstraße 25, 91054 Erlangen, Germany
e-mail: clark@chemie.uni-erlangen.de
Tel.: +49-9131-8522948, Fax: +49-9131-8526565

around the nitrogen lone-pair and the hydrogen-bond basicities of aliphatic amines. In a very recent paper, Lamarche and Platts used *ab initio* and DFT-calculations on hydrogen-bonded complexes to estimate pK_{HB}-values. [14] We are interested in the use of pK_{HB}-related descriptors to determine the strength of binding interactions and have therefore used the available pK_{HB}-data for six-membered nitrogen heterocycles [2] as the basis for a QSPR-study using, among others, descriptors derived from localized molecular orbitals (LMOs) calculated using AM1 semiempirical MO-theory. [15]

Calculational Methods

All structures were optimized without geometry constraints using the standard AM1 Hamiltonian [15] within VAMP 7.5. [16] Subsequently, a standard population analysis was performed in order to obtain the Coulson charges on the atoms [17] and localized molecular orbitals (LMOs) were calculated using the Perkins, Stewart technique. [18] The molecular electrostatic properties were calculated using the natural atomic orbital-point charge (NAO-PC) model, [19, 20] from which atomic dipoles were derived as described previously. [20] Standard van der Waals' radii [21] were used to generate the molecular surfaces. Regression analyses used TSAR 3.2. [22].

The dataset

Of the 65 compounds in the dataset, only 51 are suitable for use in the training set for several reasons. Firstly, the position at which the hydrogen bond is formed must be unique. Several compounds contain substituents, X, such as carbonyl- or cyano-groups, that form an additional 1:1 hydrogen-bond complex with 4-fluorophenol. The measured formation constant $K_f(total)$ for these compounds is a global constant corresponding to the sum of the formation constants of two 1:1 complexes:

$$K_f(total) = K_f(X) + K_f(N) \quad (3)$$

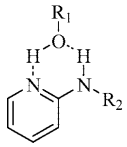
where $K_f(N)$ and $K_f(N)$ are the individual hydrogen bonding equilibrium constants for the group X and the nitrogen, respectively.

Because the individual contributions of the two formation constants are not known experimentally, these compounds cannot be used in the training set. Analogously, compounds with two or more non-equivalent ring-nitrogen atoms cannot be used because their observed pK_{HB}s are also sums of all contributions, as shown in Eq. (4).

$$K_f(total) = \sum_i K_f(N_i) \quad (4)$$

These compounds were excluded from the training-set, but can be used as a test set for the final model. A second problem arises for compounds with an amino or methylamino group *ortho* to a ring-nitrogen. In this case, 4-fluorophenol can form a doubly hydrogen-bonded structure, as shown in Scheme 2. This leads to a more strongly bound complex than would be the case with a single hydrogen bond, and hence to higher pK_{HB}-values.

Scheme 2 The *ortho*-effect in 2-amino- and 2-methylamino-pyridines



The very weak hydrogen-bond base pentafluoropyridine was also removed from the dataset, since only an uncertain pK_{HB}-value is available.

Two subsets of molecules were selected as training sets in order to develop QSPR-models for pK_{HB}. To be absolutely sure that the measured formation constant K_f is given by one well defined 1:1 hydrogen-bond complex, only molecules which contain only one possible hydrogen-bond acceptor site where chosen for the first training set. Compounds with two or three equivalent ring nitrogen atoms were also included in this training set; the experimental pK_{HB}-values for these compounds were corrected by $-\log_{10}2$ or $-\log_{10}3$, respectively. The experimental pK_{HB}-values of this subset of 42 molecules (Table 1) range from 0.14 to 2.29. In order to include the stronger hydrogen-bond bases, molecules with amino- and methoxy-substituents were added to give a second training set of 51 compounds. These groups are much weaker hydrogen-bond acceptors than the aromatic-ring nitrogen atoms, and therefore do not contribute significantly to the experimental pK_{HB}s. The experimental pK_{HB}-values for this set of molecules range from 0.14 to 2.93.

Test set

Besides the experimental pK_{HB}-values for 65 compounds, secondary pK_{HB}–values for a set of 6 compounds are given in the original experimental paper. [2] These pK_{HB}-values were calculated *via* a linear free energy relationship from the corresponding formation constants of the phenol-complex. [23] We did not use these values for training, but did include them in the test set. Additionally, compounds with two different aromatic nitrogen atoms, mentioned above, are also included in the test set. The total pK_{HB}-value for such compounds provides an additional test. Additionally, the compounds included in the larger test set, but not in the smaller one, were used as part of the test set for the smaller training set.

Descriptors

We first calculated the 17 descriptors given in Table 2 in order to test their applicability for the estimation of pK_{HB}-values.

Table 1 Calculated values for the descriptors, predicted and experimental hydrogen-bond basicities of six-membered *N*-heteroaromatic compounds. Results for compounds included in the training set are shown in boldface

Compound	q [e$^-$]	E_{LMO} [eV]	C_p	Θ [°]	pred. pK_{HB}	pred. pK_{HB}	exp. pK_{HB}
2,4,6-Trimethylpyridine	–0.149	–19.121	0.2964	73.0	2.21	2.25	2.29
Phthalazine	–0.061	–20.348	0.2547	103.1	2.29	2.31	2.27
	–0.061	–20.349	0.2543	103.0	(1.98, 2.00)[a]	(2.00, 2.02)[a]	(1.97)[b]
3,4-Dimethylpyridine	–0.140	–19.266	0.2949	101.6	2.22	2.26	2.24
3,5-Dimethylpyridine	–0.133	–19.271	0.2971	101.5	2.13	2.16	2.21
2,6-Dimethylpyridine	–0.146	–19.167	0.2970	72.9	2.08	2.11	2.14
4-*tert*-Butylpyridine	–0.141	–19.284	0.2939	101.7	2.22	2.27	2.11
4-Methylpyridine	–0.143	–19.310	0.2947	101.7	2.11	2.15	2.07
4-Ethylpyridine	–0.142	–19.309	0.2947	101.7	2.12	2.16	2.07
2-Methylpyridine	–0.142	–19.271	0.2964	73.9	1.88	1.91	2.03
3-Ethylpyridine	–0.137	–19.315	0.2961	101.5	2.06	2.09	2.01
3-Methylpyridine	–0.136	–19.318	0.2961	101.5	2.05	2.08	2.00
4-Phenylpyridine	–0.138	–19.371	0.2947	101.7	1.98	2.01	1.96
Acridine	–0.132	–19.299	0.2974	70.6	1.79	1.79	1.95
4-Vinylpyridine	–0.138	–19.360	0.2949	101.7	2.00	2.03	1.95
Isoquinoline	–0.138	–19.389	0.2951	99.7	1.90	1.92	1.94
Pyridazine	–0.047	–20.522	0.2548	101.2	1.91	1.90	1.95
	–0.046	–20.523	0.2548	101.3	(1.61, 1.61)[a]	(1.60, 1.60)[a]	(1.65)[b]
2-Ethylpyridine	–0.141	–19.250	0.2967	60.6	1.81	1.82	1.94
Quinoline	–0.130	–19.364	0.2958	70.5	1.72	1.72	1.89
2-Butylpyridine	–0.141	–19.253	0.2967	58.3	1.78	1.79	1.88
Phenanthridine	–0.129	–19.387	0.2964	68.0	1.61	1.61	1.87
Pyridine	–0.139	–19.364	0.2953	101.6	1.96	1.99	1.86
2-Isopropylpyridine	–0.139	–19.225	0.2975	54.3	1.79	1.80	1.76
2-Vinylpyridine	–0.140	–19.276	0.2976	34.6	1.48	1.47	1.65
4-Chloropyridine	–0.136	–19.621	0.2942	101.6	1.40	1.40	1.54
Phenazine	–0.095	–19.663	0.2971	71.7	1.41	1.35	1.52
	–0.095	–19.664	0.2971	71.7	(1.11, 1.10)[a]	(1.05, 1.05)[a]	(1.22)[b]
2,2'-Bipyridine	–0.133	–19.389	0.2969	57.2	1.77	1.76	1.45
	–0.133	–19.390	0.2969	57.3	(1.47, 1.47)[a]	(1.46, 1.46)[a]	(1.15)[b]
2-Phenylpyridine	–0.133	–19.339	0.2952	57.8	1.68	1.69	1.43
2-*tert*-Butylpyridine	–0.140	–19.181	0.2987	44.2	1.74	1.75	1.42
Pyrimidine	–0.165	–19.937	0.2809	101.3	1.43	1.46	1.37
	–0.165	–19.939	0.2806	101.4	(1.12, 1.13)[a]	(1.15, 1.17)[a]	(1.07)[b]
3-Iodopyridine	–0.138	–19.679	0.2932	101.3	1.29	1.29	1.37
3-Fluoropyridine	–0.115	–19.678	0.2958	101.0	1.28	1.26	1.35
3-Chloropyridine	–0.128	–19.653	0.2940	101.3	1.37	1.36	1.31
3-Bromopyridine	–0.135	–19.692	0.2940	101.3	1.24	1.23	1.31
Pyrazine	–0.102	–19.832	0.2934	102.4	1.40	1.36	1.22
	–0.102	–19.833	0.2934	102.4	(1.10, 1.10)[a]	(1.06, 1.06)[a]	(0.92)[b]
7,8-Benzoquinoline	–0.139	–19.344	0.2991	26.4	1.17	1.14	1.16
2-Chloropyridine	–0.128	–19.676	0.2927	83.3	1.22	1.20	1.05
2-Fluoropyridine	–0.159	–19.779	0.2897	93.0	1.03	1.04	0.95
2-Bromopyridine	–0.113	–19.710	0.2932	78.6	1.15	1.12	1.03
5-Bromopyrimidine	–0.162	–20.240	0.2796	101.3	0.75	0.75	0.89
	–0.162	–20.241	0.2795	101.3	(0.45, 0.45)[a]	(0.45, 0.45)[a]	(0.59)[b]
3,5-Dichloropyridine	–0.118	–19.922	0.2927	101.3	0.82	0.78	0.85
s-Triazine	–0.190	–20.551	0.2638	103.4	0.81	0.85	0.80
	–0.190	–20.553	0.2638	103.4	(0.34, 0.33, 0.32)[a]	(0.38, 0.37, 0.36)[a]	(0.32)[c]
	–0.191	–20.549	0.2642	103.4			
2,6-Difluoropyridine	–0.188	–20.131	0.2869	92.4	0.14	0.14	0.14
4-Pyrrolidinopyridine	–0.171	–19.115	0.2897	101.8	2.69	2.79	2.93
4-*N*,*N*-Diethylaminopyridine	–0.178	–19.054	0.2884	101.9	2.87	2.98	2.89
4-*N*,*N*-Dimethylaminopyridine	–0.167	–19.165	0.2904	101.8	2.55	2.64	2.80
4-(4-Methylpiperidino)pyridine	–0.163	–19.177	0.2914	101.8	2.49	2.57	2.68
4-Piperidinopyridine	–0.163	–19.181	0.2909	101.8	2.51	2.59	2.68
4-Aminopyridine	–0.172	–19.188	0.2895	101.7	2.52	2.61	2.56
3-Aminopyridine	–0.109	–19.219	0.3000	101.1	2.24	2.25	2.20
4-Methoxypyridine	–0.162	–19.356	0.2915	101.5	2.05	2.10	2.13
2-*N*,*N*-Dimethylaminopyridine	–0.212	–19.139	0.2936	25.9	1.56	1.62	1.61
1,7-Phenanthroline	–0.14	–19.501	0.2988	26.6	1.64	1.64	1.87
	–0.138	–19.425	0.2946	69.6	(0.78, 1.58)[a]	(0.74, 1.58)[a]	
Quinazoline	–0.152	–19.831	0.2807	73.7	1.49	1.52	1.55
	–0.170	–19.930	0.2799	99.2	(1.23, 1.15)[a]	(1.25, 1.18)[a]	
3-(*N*,*N*-Dimethylamino)pyridine	–0.122	–19.235	0.2988	101.1	2.19	2.21	2.43[d]

Table 1 (continued)

Compound	q [e$^-$]	E_{LMO} [eV]	C_p	Θ [°]	pred. pK_{HB}	pred. pK_{HB}	exp. pK_{HB}
2-(*N*,*N*-Dimethylamino)pyrimidine	–0.235	–19.636	0.2768	23.1	1.52	1.61	
	–0.232	–19.627	0.2749	46.3	(1.02, 1.36)[a]	(1.10, 1.46)[a]	1.40[d]
4-(*N*,*N*-Dimethylamino)pyrimidine	–0.229	–19.666	0.2785	44.8	2.05	2.16	2.11[d]
	–0.204	–19.614	0.2752	101.6	(1.08, 2.01)[a]	(1.16, 2.12)[a]	
5-(*N*,*N*-Dimethylamino)pyrimidine	–0.146	–19.750	0.2856	101.1	1.76	1.79	1.88[d]
	–0.143	–19.747	0.2856	101.2	(1.45, 1.48)[a]	(1.47, 1.50)[a]	
4-(*N*,*N*-Dimethylamino)quinoline	–0.146	–19.256	0.2929	69.5	2.04	2.07	2.43[d]
9-(*N*,*N*-Dimethylamino)acridine	–0.147	–19.107	0.2957	69.4	2.26	2.30	2.31[d]

[a] The values in brackets are the calculated $-\log_{10}K_f$ values for the individual hydrogen-bond acceptor sites. The net pK_{HB}-values were calculated using Eq. (4).
[b] The values in brackets are corrected statistically by $-\log_{10}2$.
[c] The values in brackets are corrected statistically by $-\log_{10}3$.
[d] Calculated from the equation: pK_{HB} =1.198 log K–0.126. K is the formation constant of the phenol-pyridine complex at 298 K in CCl_4.

Table 2 Calculated descriptors

q	Coulson charge on the nitrogen atom
q_{Mull}	Mulliken charge on the nitrogen atom
q_{ESP}	VESPA charge on the nitrogen atom [24]
μ_N	Atomic dipole moment of the nitrogen atom [20]
Q_N	Atomic quadrupole moment of the nitrogen atom [20]
α_N	Atomic polarizability [26]
V_N	Atomic valence of the nitrogen atom [18]
E_{HOMO-1}	Eigenvalue of the HOMO-1
E_{HOMO}	Eigenvalue of the HOMO
E_{LUMO}	Eigenvalue of the LUMO
E_{LUMO+1}	Eigenvalue of the LUMO+1
E_{LMO}	Eigenvalue of the localized nitrogen lone pair orbital [18]
C_p	Coefficient of the *p*-orbital in the nitrogen lone pair orbital [18]
spX	Hybridzation of the nitrogen lone pair orbital [18]
Θ	Access angle (see text)
sin Θ	Sin of the access angle
ESP_{min}	Minimum electrostatic potential at the molecular surface

Our final model was obtained by selection from these descriptors using individual correlation coefficients with the target property with subsequent tests of up to six-descriptor models. It uses four descriptors whose direct relationship to pK_{HB} is easily visualized. These are:

Coulson charge, q. The partial charge of the nitrogen atom is expected to play a major role because hydrogen bonds are predominantly electrostatic in nature. [3, 4, 5] In principle, many different types of partial charge could be used here, although they fall into two classes, those derived from an analysis of the electron density and those derived from the molecular electrostatic potential (MEP). We have chosen to use the Coulson charges [17] because the alternative Mulliken charges gave almost identical results and the more computationally expensive potential-derived charges calculated with our VESPA technique [24] gave a slightly worse model. The Coulson charges were therefore chosen for simplicity and computational efficiency.

Energy of the localized lone pair, E_{LMO}. The covalent component of a hydrogen bond [4] can be treated as a donor-acceptor interaction in which the σ^*_{OH}-orbital of the H-bond donor acts as the accepting orbital (see Fig. 1). This interaction depends on the energy gap between this σ^*_{OH}-orbital, which is constant in pK_{HB}-measurements, and the lone pair on the H-bond acceptor nitrogen atom. [3, 5] The ionization potential of the lone pair is known to vary inversely with both the charge-transfer within a hydrogen bond and the charge-redistribution within the hydrogen bond acceptor. [5] In contrast to earlier work, however, we have used the localized orbital energy in order to be able to identify the lone pair clearly. The canonical molecular orbitals corresponding to in plane lone pairs are often considerably delocalized.

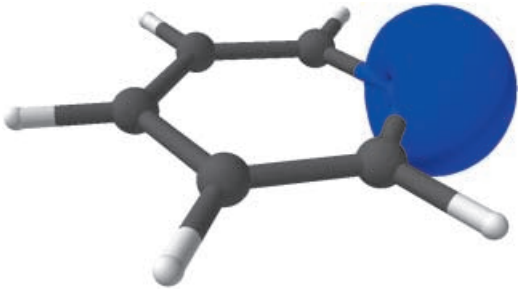


Fig. 1 Localized molecular orbital of the lone pair in pyridine

Fraction of p-character in the localized lone pair, C_p. Although the C_p-character in the localized lone pair is fairly highly correlated with the LMO-energy (r^2=0.73), it contains extra information about the inductive effect of

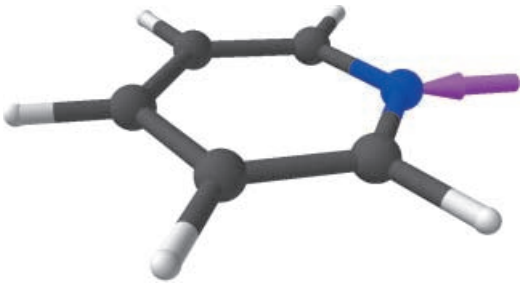
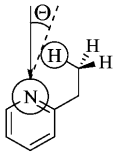


Fig. 2 Atomic dipole moment of the nitrogen atom in pyridine

Scheme 3 Calculation of the access angle Θ



the neighboring atoms and about the geometry around the nitrogen atom in question (see Fig. 1).

Access angle, Θ. The smallest angle between the calculated atomic dipole vector for the nitrogen atom in question (Fig. 2) and the van der Waals' surface of the molecule, as defined in Scheme 3, was used to describe the steric hindrance around the acceptor site. We considered the alternative of taking the access angle from the calculated geometry of the hydrogen-bonded complex, but this would result in a considerably larger computational task and would have interfered with our objective of setting up a model derived only from the calculated properties of the hydrogen-bond acceptor.

Results

1. Small training set

The results obtained for the small (42 compound) training set are shown in Fig. 3.

The regression equation obtained for this dataset is:

$$\begin{aligned} pK_{HB} &= 5.335q + 2.478E_{LMO} - 50.051C_p \\ &\quad + 0.008687\Theta + 64.592 \\ r &= 0.9707, r^2 = 0.9423, r^2(CV) = 0.9303, \\ F &= 5.96 \cdot 10^{-24}, \sigma = 0.136 \end{aligned} \tag{5}$$

where F is the F-probability and σ is the standard error of the regression model. The regression coefficients obtained for the data scaled and standardized by their mean and standard deviations are 0.134, 0.922, 0.520 and 0.187 for q, E_{LMO}, C_p and Θ, respectively. These standardized coefficients indicate the relative importance of the four descriptors without the effects of their very different magnitudes and ranges and indicate that each descriptor plays a significant role in the regression equation. Table 3 shows the cross-correlation matrix for the four descriptors. Only the lone-pair energy and the *p*-contribution show a significant correlation, but this is below the 90% threshold that we usually use to eliminate correlated descriptors in regression models.

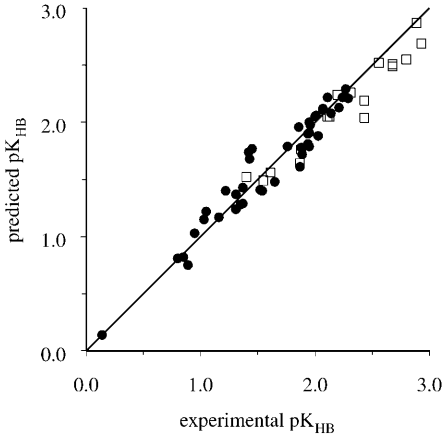


Fig. 3 Predicted vs. experimental pK_{HB}-values for the small (42 compound) training set. Compounds from the training set are shown as *closed circles*, those from the test set as *open squares*

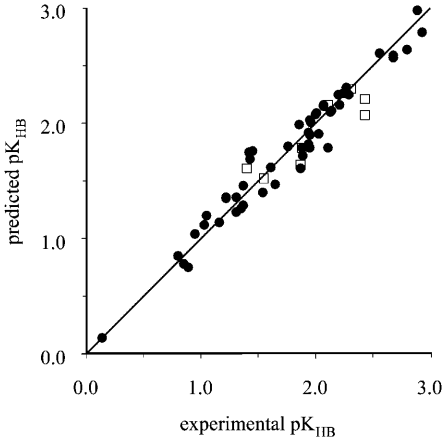


Fig. 4 Predicted vs experimental pK_{HB}-values for the large (51 compound) training set. Compounds from the training set are shown as *closed circles*, those from the test set as *open squares*

The mean unsigned error for the nine well-defined molecules from the test set (i.e. those that were included in the second training set) is 0.12 log units and for the complete test set of 17 compounds 0.14 log units. The mean unsigned error for the eight additional compounds is 0.16 log units. The largest errors are –0.24 and –0.39 log units for the small and large test sets, respectively. Encouragingly, the calculated pK_{HB}s for the test set molecules are systematically too low, as they should be if other sites are contributing to the experimental binding constant.

2. Large training set

The results obtained for the 51 compound training set are shown in Fig. 4.

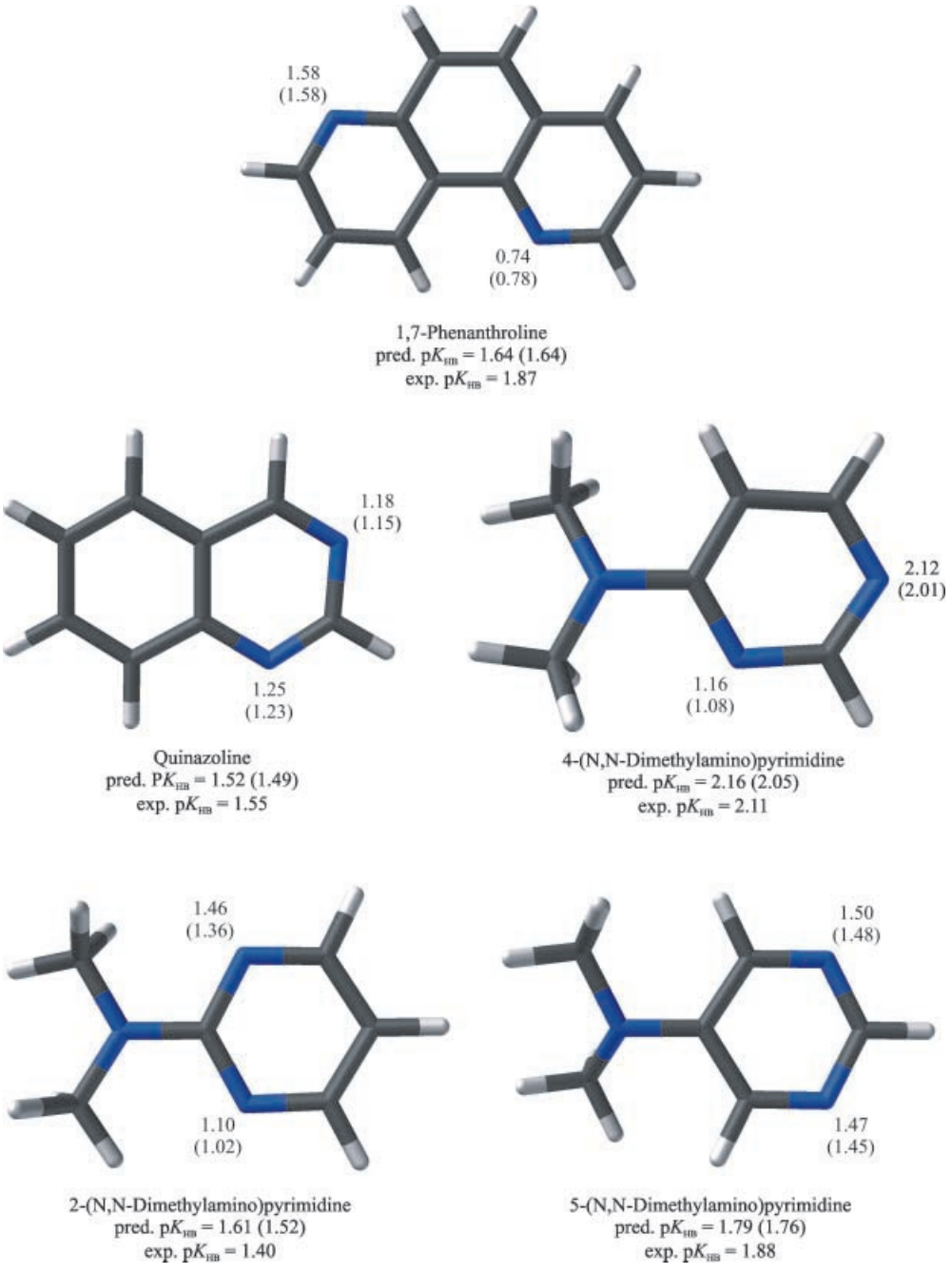


Fig. 5 Predicted $\log_{10}K_f$ and pK_{HB}-values for compounds with two aromatic nitrogen atoms, calculated using regression Eq. (7). The values in brackets are calculated using regression Eq. (6)

The regression equation obtained is:

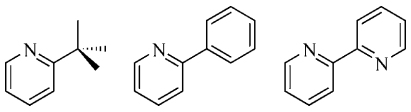
$$\begin{aligned} pK_{HB} &= 4.724q + 2.593E_{LMO} - 53.734C_p \\ &\quad + 0.009255\Theta + 67.786 \\ r &= 0.9799,\ r^2 = 0.9602,\ r^2\,(CV) = 0.9530, \\ F &= 1.88 \cdot 10^{-33},\ \sigma = 0.131 \end{aligned} \tag{6}$$

The mean unsigned and largest errors for the eight compound test set are 0.12 and 0.36 log units, respectively. The largest error is given by both regression equations for 4-(*N, N*-dimethylamino)quinoline. Figure 5 shows the calculated pK_{HB}-values for the five compounds from the test set with two ring nitrogens calculated with the two regression equations. These compounds provide essentially a blind test because the individual pK_{HB}-values are unknown.

Table 3 Correlation coefficients *r* between pK_{HB} and the four descriptors *q*, E_{LMO}, C_p and Θ

	pK_{HB}	*q*	E_{LMO}	C_p	Θ
pK_{HB}	1.00				
q	0.24	1.00			
E_{LMO}	0.68	–0.20	1.00		
C_p	0.29	–0.32	0.85	1.00	
Θ	–0.03	0.06	–0.45	–0.37	1.00

The two regression equations are pleasingly similar, suggesting that the additional compounds included in the



Scheme 4 Sterically crowded compounds: 2-*tert*-butylpyridine, 2-phenylpyridine and 2,2'-bipyridine

large training set and not in the smaller, more conservative, one do not have any adverse effects on the correlation.

Both regression equations predict pK_{HB}-values about 0.3 log units too high for the three sterically crowded compounds shown in Scheme 4.

These compounds might be expected to lose some rotational freedom of the substituent on formation of a hydrogen bond to the ring nitrogen. Therefore a contribution from rotatable groups in *ortho*-position (ΔG_{rot}) should be taken into account. According to Böhm, this contribution is about 1.4 kJ mol^{-1}, [25] which corresponds to –0.25 pK_{HB}-units. We have therefore used the number of rotatable groups (N_{rot}) at the *ortho*-positions as an additional descriptor. This includes several compounds with ethyl-, n-butyl- and iso-propyl-groups in *ortho*-position, although the influence of these groups seems to be lower. The regression equation obtained is:

$$\begin{aligned} pK_{HB} &= -0.1889N_{rot} + 4.951q + 2.614E_{LMO} \\ &\quad - 54.020\%p + 0.007867 + 68.444 \\ r &= 0.9834,\ r^2 = 0.9670,\ r^2(CV) = 0.9567, \\ f &= 263.9,\ \sigma = 0.121 \end{aligned} \qquad (7)$$

The resulting regression equation has r^2_{cv} = 0.96, which is only a slight improvement, but the pK_{HB}-values of the three compounds mentioned above are reproduced more accurately. The regression coefficient for N_{rot} of –0.19 is in reasonable agreement with the values given by Böhm (–0.24). [25]

Discussion

The success of the simple QSPR-models presented here is impressive, especially as calculating the hydrogen-bonding energies of complexes of the bases with 4-fluorophenol directly with AM1 leads to no correlation at all between pK_{HB} and the calculated complexation energies. Thus, the black magic of QSPR-regressions can extract a useful and predictive correlation from a calculational method that, when used directly, does not reproduce the effects being studied.

The present study is necessarily limited by the available data and should be understood as an investigation of the feasibility of calculating hydrogen bond acidities for applications such as QSAR and scoring functions for docking techniques. In this respect, we believe that we have shown that hydrogen-bonding ability can be quantified with few, relatively simple descriptors. The regression equations given here are only applicable to very similar compounds to those used here (i.e. to six-membered ring nitrogen bases with sp^2-lone pairs) and will not perform properly for, for instance, amines with sp^3-lone pairs. Nevertheless, the importance of two descriptors that describe the localized nitrogen lone pair is an important indication that LM*O*-properties can be useful descriptors for QSPR-studies involving specific, local interactions, such as pK_{HB} or pK_A. The use of LM*O*-properties can be understood as an expression of the fact that hydrogen bond formation or protonation perturb the lone pair in a similar way to the mathematical localization process.

The present study provides the basis for more general 3D-descriptors designed to quantify both the direction and the strength of hydrogen bond donors and acceptors.

Acknowledgements This work was supported by the Fonds der Chemischen Industrie and the Deutsche Forschungsgemeinschaft.

References

1. Laurence C, Berthelot M (2000) Perspectives in Drug Discovery and Design 18:39–60
2. Berthelot M, Laurence C, Safar M, Besseau F (1998) J Chem Soc, Perkin Trans 2:283–290
3. Gordon MS, Jensen JH (1996) Accts Chem Res 29:536–543
4. Kollman P, McKelvey J, Johansson A, Rothenberg S (1975) J Am Chem Soc 97:955–965
5. Kollman PA, Allen LC (1971) J Am Chem Soc 93:4991–5000
6. Murray JS, Politzer P (1991) J Org Chem 56:3734–3737
7. Murray JS, Politzer P (1991) J Org Chem 56:6715–6717
8. Hagelin H, Murray JS, Brinck T, Berthelot M, Politzer P (1995) Can J Chem 73:483–488
9. Kenny PW (1994) J Chem Soc, Perkin Trans 2:199–202
10. Le Questel JY, Berthelot M, Laurence C (1997) J Chem Soc, Perkin Trans 2:2711–2717
11. Herail M, Berthelot M, Proutiére A (1995) J Phys Org Chem 8:421–428
12. Herail M, Magnassan E, Proutiére A (1997) J Phys Org Chem 10:167–174
13. Graton J, Laurence C, Berthelot M, Le Questel JY, Besseau F, Raczynska EW (1999) J Chem Soc, Perkin Trans 2:997–1001
14. Lamarche O, Platts JA (2002) Chem Eur J 8:457–466
15. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909; Holder AJ (1998) AM1. In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer III HF, Schreiner PR (eds) Encyclopedia of Computational Chemistry, 1st edn, vol 1. Wiley, Chichester, pp 8–11
16. Clark T, Alex A, Beck B, Chandrasekhar J, Gedeck P, Horn A, Hutter M, Martin B, Rauhut G, Sauer W, Schindler T, Steinke T (2000) Vamp 7.5. Oxford Molecular, The Medawar Centre, Oxford Science Park, Sandford–on–Thames, Oxford OX4 4GA, United Kingdom
17. Coulson CA, Longuet-Higgins HC (1947) Proc Roy Soc (London) A191:39–60
18. Perkins PG, Stewart JJP (1982) J Chem Soc, Faraday Trans 2 78:285–296
19. Rauhut G, Clark T (1993) J Comput Chem 14:503–509
20. Beck B, Rauhut G, Clark T (1994) J Comput Chem 15:1064–1073
21. Bondi A (1964) J Phys Chem 68:441–451
22. Oxford Molecular (1998) TSAR 3.2. Oxford Molecular, The Medawar Centre, Oxford Science Park, Sandford–on–Thames, Oxford OX4 4GA, United Kingdom
23. Cruège F, Girault G, Coustal S, Lascombe J, Rumpf P (1970) Bull Soc Chim Fr 11:3889–3894
24. Beck B, Clark T, Glen, RC (1995) J Mol Model 1:176–187
25. Böhm HJ (1994) J Comput-Aided Mol Des 8:243–256
26. Martin B, Gedeck P, Clark T (2000) Int J Quant Chem 77:473–497