Journal of Computer-Aided Molecular Design, **16:** 569–583, 2002.
KLUWER/ESCOM

Fast estimation of hydrogen-bonding donor and acceptor propensities: a GMIPp study

Albert Salichs[1], M. López[2], V. Segarra[2], Modesto Orozco[3]* & F. Javier Luque[1]*
[1]*Departament de Fisicoquímica. Facultat de Farmàcia. Universitat de Barcelona. Avgda. Diagonal s/n. Barcelona 08028. Spain;* [2]*Almirall-Prodesfarma Research Centre, c/ Cardener 68-74, 08024 Barcelona, Spain;* [3]*Departament de Bioquímica i Biologia Molecular. Facultat de Química. Universitat de Barcelona. c/. Martí Franquès 1. 08028 Barcelona. Spain*

Received 10 January 2001; accepted in final form 13 October 2002

Key words: hydrogen-bonding, molecular interaction potential, quantum mechanical properties, hydrogen-bond basicity, hydrogen-bond acidity, electrostatic potential

Summary

The suitability of the GMIPp energy functional as a fast, efficient method for estimating the hydrogen-bond donor and acceptor propensities of a wide variety of organic compounds is examined. Comparison of the GMIPp values is made with two experimental hydrogen-bond scales: i) the hydrogen-bond basicity scale for N-heteroaromatics in carbon tetrachloride, and ii) the hydrogen-bond acidities for NH/OH donors and hydrogen-bond basicities of N/O acceptors determined in 1,1,1-trichloroethane. Attention is paid to i) the reliability of semiempirical *versus ab initio* treatments of the quantum mechanical molecule, ii) the role of solvation, and iii) the effect of including the polarization energy component in the calculation of the GMIPp functional. The statistical analysis of the results reveals that the GMIP functional, which combines electrostatic and steric energy components, predicts with reasonable accuracy and computational efficiency the hydrogen-bond strength for a wide variety of compounds.

Introduction

Hydrogen bonding plays a key role in modulating intermolecular interactions in chemical and biological systems [1–3]. This weak interaction occurs when a donor group, typically a hydrogen bound to an heteroatom, is close to an electron-rich acceptor group, such as an heteroatom or the π-electron distribution of aromatic rings. Hydrogen bonding is often a key interaction in the formation of non-covalent *host-guest* complexes, which explains its fundamental role in supramolecular chemistry [4, 5]. Furthermore, hydrogen bonds are responsible of the stabilization of proteins and DNA structures, the maintenance of the genetic code, the specific recognition of ligands by macromolecules and many other biological phenomena [6].

*Correspondence to F. J. Luque (javier @ far1.far.ub.es) or M. Orozco (modesto @ luz.bq.ub.es).

High-level quantum mechanical (QM) methods can provide accurate geometrical, spectroscopic and energetic features of hydrogen bonded complexes [1–3] Moreover, they allow us to distinguish the nature of the forces that stabilize hydrogen-bonded complexes [7–10], the role of secondary interactions in multiple hydrogen-bonded complexes [11], or even to examine the functional implications of proton transfer dynamics in hydrogen-bonded complexes [12–14]. The use of high-level QM methods is, nevertheless, severely limited by their computational expensiveness, which precludes their application in the study of hydrogen-bonding properties of large molecules or in the analysis of very large data bases of small and medium sized molecules.

A qualitative description of the hydrogen-bonding capabilities of molecules is valuable in certain research areas. Thus, simple counting of hydrogen bond donors and acceptors is used in the prediction of solubility and transport properties of drugs, in the

design of new ligands, and in the structural analysis of macromolecules [15–18]. This approach is also used to screen the pharmacokinetic properties of compounds in real or virtual libraries of compounds [19, 20]. Such a qualitative description has, however, several drawbacks. First, there may be differences in the counting of hydrogen-bond donors and acceptors owing to the use of different empirical rules [21, 22]. Moreover, such counting approach neglects the effect of neighboring functional groups, which might change the hydrogen-bond donor/acceptor properties [23] Finally, the influence of factors like molecular conformation, which can limit the accessibility of a given donor/acceptor group, is omitted.

In order to seek for more precise definitions of hydrogen-bonding capabilities, attempts have been made to derive more accurate experimental hydrogen-bonding parameters. These attempts are based on the analysis of solvatochromic data [24–26], or alternatively derived from the comparison between octanol-water and cyclohexane-water partition coefficients [27, 28] or from the difference in octanol-water logP of polar and apolar compounds of the same size [29]. Besides the usefulness of these empirical techniques, some disadvantages are evident. Thus, as the hydrogen parameters might be influenced by the experimental procedure, their transferability is not fully guaranteed. Furthermore, there can be difficulties in determining the hydrogen-bonding parameters for certain types of chemical structures under specific experimental conditions. Finally, experimental values may be missing for particular chemical groups.

Recently, attempts have been made to overcome these difficulties by resorting to QM-derived parameters, like the charge on the hydrogen-bonding hydrogen or acceptor atoms, the energy of the lowest unoccupied/highest occupied molecular orbitals, the electron donor superdelocalizability, the self-atom polarizability, or the molecular electrostatic potential [30–34]. To further explore the capability of theoretical methods to define hydrogen-bonding scales, this study examines the suitability of the *Generalized Molecular Interaction Potential with Polarization* (GMIPp; [35–37]) to explain the hydrogen-bonding ability of molecules. This mixed QM/MM parameter is determined with a computational cost similar to that of the MEP, and has been shown extremely useful to reproduce many non-covalent interactions [38–43], which suggests that it can be useful to explore the hydrogen-bonding donor and acceptor properties of compounds.

Methods

The GMIPp functional

The QM interaction energy between two molecules can be decomposed into energy components following different partitioning schemes [44–46]. Generally, the interaction energy can be decomposed into SCF and dispersion contributions. Moreover, the SCF interaction energy can be divided into i) electrostatic, ii) polarization, iii) charge transfer, and iv) exchange. For most weak non-covalent complexes involving polar compounds the electrostatic and polarization contributions are the leading terms of the SCF interaction energy. This justifies the success of QM/MM methods [47, 48], where one of the interacting particles is treated classically, and accordingly exchange and charge transfer terms are taken into account implicitely.

The electrostatic interaction energy between a QM molecule and a classical system can be determined from the molecular electrostatic potential [49] created the QM molecule, V_{QM}, as noted in Equation 1.

$$E_{ele} = \sum_{k} Q_k V_{QM}(\mathbf{r_k}) = \sum_{k}\sum_{A} \frac{Q_k Z_A}{|\mathbf{r_k} - \mathbf{R_A}|} - \sum_{k}\sum_{i}^{occ}\sum_{\mu}\sum_{\nu} c_{\mu i} c_{\nu i} \left\langle \phi_\mu \left| \frac{Q_k}{|\mathbf{r_k} - \mathbf{r}|} \right| \phi_\nu \right\rangle \quad (1)$$

where Q_k denotes the set of *k* point charges placed at $\mathbf{r}_k$ that describe the charge distribution of the classical particle, Z_A is the nuclear charge of atom A, placed at $\mathbf{R_A}$, ϕ stands for the set of basis functions used for the QM molecule, and $c_{\mu i}$ is the coefficient of the atomic orbital μ in the molecular orbital i. At the Hartree-Fock or DFT level of theory, the summation over the molecular orbitals is limited to the doubly occupied orbitals for a closed-shell molecule.

Polarization effects can be included in a computationally efficient way by using a perturbative expression derived by Francl [50]. At the Hartree-Fock level of theory, the interaction of the QM molecule with the set of point charges of the classical particle is given by Equation 2.

$$E_{pol} = \sum_{i}^{occ}\sum_{j}^{vir} \frac{1}{\varepsilon_i - \varepsilon_j}$$

$$\left\{\sum_{\mu}\sum_{\nu} c_{\mu i}c_{\nu i}\left\langle\phi_{\mu}\left|\frac{Q_k}{|\mathbf{r_k}-\mathbf{r}|}\right|\phi_{\nu}\right\rangle\right\}^2 \quad (2)$$

where the second summation in the right-hand side runs over virtual, unoccupied molecular orbital, and ε_i denotes the energy of molecular orbital *i*,

Dispersion and repulsion contributions to the interaction energy is typically accounted for by using a pairwise-additive 6–12 Lennard-Jones potential (Equation 3).

$$E_{ste} = \sum_{k}\sum_{A}\left(\frac{C_{kA}}{|\mathbf{r_k}-\mathbf{R_A}|^{12}} - \frac{D_{kA}}{|\mathbf{r_k}-\mathbf{R_A}|^{6}}\right) \quad (3)$$

where C_{kA} and D_{kA} are empirical parameters for the repulsive and dispersion components, which uses to be computed by means of combinatorial rules based on atomic radius (r^*) and hardness (ε^*), as noted in Equations 4a–c.

$$E_{ste} = \sum_{k}\sum_{A}\varepsilon^*_{kA}\left[\left(\frac{r^*_{kA}}{|\mathbf{r_k}-\mathbf{R_A}|}\right)^{12} - 2\left(\frac{r^*_{kA}}{|\mathbf{r_k}-\mathbf{R_A}|}\right)^{6}\right] \quad (4a)$$

$$\varepsilon^*_{kA} = (\varepsilon^*_k\varepsilon^*_A) \quad (4b)$$

$$r^*_{kA} = r^*_k + r^*_A \quad (4c)$$

The addition of electrostatic (Equation 1) and steric (Equation 4) energy terms allowed us to define the *Generalized Molecular Interaction Potential* (GMIP), whereas further addition of the polarization (Equation 2) term was denoted as the *Generalized Molecular Interaction Potential with polarization* (GMIPp) [35–37]. Inspection of Equations 1–4 shows that the method gives estimates of the interaction energy between a QM molecule and a classical system placed at a given position in the space avoiding the recalculation of the wavefunction of the QM particle. In fact, only monoelectron integrals need to be recomputed, which leads to a very efficient computational method [37, 43].

The MST method

The influence of the solvent on the intrinsic hydrogen-bonding properties has been examined from the wavefunction of the compounds in solution determined by using the MST continuum model [51–54], also known as Polarizable Continuum model [55]. The MST method computes the free energy of solvation as the sum of cavitation, van der Waals, and electrostatic contributions.

The cavitation free energy contribution is computed following Pierotti-Claverie scaled particle theory [56, 57], and the van der Waals term is determined using an empirically-adjusted linear relationship with the atomic solvent-exposed surface [51–54]. The electrostatic term is determined treating the solvent as a continuous polarizable medium, which reacts against the solute charge distribution. The reaction field generated by the solvent is introduced as a perturbation operator, V_R, into the Schrödinger equation (Equation 5). The perturbation operator itself is expressed in terms of a set of imaginary charges spread over the solute cavity (Equation 6) by solving the Laplace equation (Equation 7) with suitable boundary conditions [55, 58]. Owing to the mutual dependence between the solute wavefunction and the solvent reaction field, Equations 5–7 must be solved until self-consistency is achieved.

$$(H^o + V_R)\Psi = E\Psi \quad (5)$$

$$V_R = \sum_{i=1}^{M}\frac{q_j}{|r_j - r|} \quad (6)$$

where M is the total number of surface elements, j, into which the solute/solvent boundary is divided, and $\{q_j\}$ is the set of charges (located at r_j) that represents the solvent response.

$$q_j = -\frac{\varepsilon - 1}{4\pi\varepsilon}S_j\left(\frac{\partial V_T}{\partial n}\right)_j \quad (7)$$

where V_T is the total electrostatic potential, which includes both solute and solvent contributions, n is the unit vector normal to the surface element j, S_j is the area of the surface element j, and ε is the solvent dielectric constant.

Solving the nonlinear Schrödinger equation (Equation 5) provides the electrostatic contribution to the free energy of solvation (Equation 8) and the fully relaxed electronic wavefunction of the solute in solution, which allows us to compare the results obtained from the solvent-adapted (Ψ) and gas phase (Ψ^0) wavefunctions

$$\Delta G_{ele} = \left\langle\Psi^{sol}\left|H^o + \frac{1}{2}V^{sol}\right|\Psi^{sol}\right\rangle - \left\langle\Psi^o|H^o|\Psi^o\right\rangle \quad (8)$$

where the perturbation operator V^{sol} accounts for the reaction field generated by the solvent in presence of the fully polarized solute.

Computational details

Molecular geometries of the compounds were fully optimized in the gas phase using the semiempirical AM1 hamiltonian [59] and *ab initio* Hartree-Fock methods with the 6-31G(d) [60] basis set. These calculations were performed using a locally modified version of MOPAC6.0 [61] and Gaussian 94 [62], respectively. GMIPp calculations were performed using the gas phase wavefunction determined at the HF level. The electrostatic potential at the semiempirical AM1 level was determined by using the NDDO-based strategy [63, 64]. Following our original model [35], van der Waals parameters taken from the TRIPOS/5 force field [65] were used. For the classical particle, a charge of $+0.5/-0.5$ (in units of electron) and van der Waals parameters of hydrogen/oxygen atoms for the study of hydrogen-bond basicity and acidity, respectively, were used. GMIPp calculations were performed by using our MOPETE [66] computer program. The effect of solvation was investigated from the wavefunction of the compounds in solution determined by using the AM1-optimized version of the MST model [67] implemented in a locally modified version of MOPAC6.0 [61].

Results and discussion

To investigate the suitability of the GMIPp functional to predict hydrogen-bonding strength, we have used the experimental hydrogen-bonding scales reported by Berthelot et al. [68] and by Abraham et al. [69]. Berthelot et al. reported a hydrogen-bond acceptor scale for N-heteroaromatics from the equilibrium constants for the formation of 1:1 hydrogen bonded complexes with 4-fluorophenol in carbon tetrachloride. Similarly, Abraham and coworkers determined hydrogen-bonding donor and acceptor scales for a large series of donor and acceptors compounds, including diverse functional groups. Hydrogen bonding equilibrium constants were measured in 1,1,1-trichloroethane using N-methylpyrrolidine and 4-nitrophenol as reference hydrogen-bond acceptor and donors.

In order to obtain fast estimates of hydrogen-bond strength, MIP calculations were performed treating the QM molecule at the AM1 semiempirical level. In some cases, calculations were also performed treating the hydrogen-bond donor/acceptor at the HF/6-31G(d) level. Selected calculations were carried out using solvent-relaxed wavefunction determined from MST calculations, which allowed us to explore the influence of solvation on the intrinsic hydrogen-bonding properties. Finally, the effect of polarization on hydrogen-bonding was considered by means of MIPp computations in selected cases.

pK_β of N-heterocycles

Table 1 reports the pK_β values of the pyridine-like nitrogen atom determined for 61 N-heteroaromatic compounds by Berthelot et al. [68]. The series covers a wide range of pK_β values varying from 2.93 to -0.49 corresponding to pyridines diversely substituted at the *ortho*, *meta* and *para* positions, pyridines with one or two fused benzene rings, diazines, and *s*-triazine. Table 1 also gives the energy values corresponding to the interaction between the QM compound and the classical probe determined from MIP calculations performed by using the AM1 wavefunction of the compound in the gas phase.

With exception of five compounds (2-fluoro, 2-chloro, 2-cyano, 2,6-difluoro and pentafluoropyridine), there is a linear relationship between the pK_β values and the MIP energies, as noted in Equation 9, which explains near 85% of the variance in the pK_β values for the N-heteroaromatic compounds. The exclusion of the *ortho*-substituted halogen/cyano derivatives from the regression line was motivated to the overestimated depth (in absolute value) of the semiempirical AM1 electrostatic potential minimum around the pyridine-like nitrogen compared to the ab initio HF/6-31G(d) value. However, such a deviation might be corrected by scaling the electrostatic component of the MIP, as will be discussed below.

$$pK_\beta = -0.25(0.01)MIP_{AM1} - 2.58(0.25)$$
$$r = 0.92 \quad r^2_{cv} = 0.84 \quad n = 56 \quad F = 309.8 \qquad (9)$$

where the standard error of the fitted coefficients is given in parenthesis, r is the Pearson's correlation coefficient, r^2_{cv} is the leave-one-out cross-validated coefficient of the regression, n is the number of data, and F is Fischer's F-statistics.

We also determined the MIP using the wavefunction of the solute in carbon tetrachloride from MST/AM1 calculations. As expected from the low dielectric permittivity of carbon tetrachloride, the MIP

Table 1. Hydrogen-bond basicity[a] (pK_β) of N-heteroaromatic compounds and the interaction energy (kcal/mol) determined from MIP calculations by using the AM1 wavefunction *in vacuo* (MIP) or in carbon tetrachloride solution (MIP_{CCl4}) and from the HF/6-31G(d) wavefunction without ($MIP_{HF/6-31G(d)}$) or with (MIPp) polarization energy.

No.	Compound	pK_β	MIP_{AM1}	MIP_{CCl4}	$MIP_{HF/6-31G*}$	MIPp
1	4-Pyrrolidinopyridine	2.93	−21.8	−22.6		
2	4-N,N-Diethylaminopyridine	2.89	−21.1	−21.9		
3	4-N,N-Dimethylaminopyridine	2.80	−20.8	−21.5		
4	4-(4-Methylpiperidino)pyridine	2.68	−21.3	−22.2		
5	4-Piperidinopyridine	2.68	−21.1	−21.7		
6	N-Methyl-N-pyridin-4-ylhydrazine	2.58	−21.5	−22.7		
7	4-Aminopyridine	2.56	−20.2	−20.9	−23.1	−30.2
8	2,4,6-Trimethylpyridine	2.29	−19.8	−20.5		
9	Phthalazine	2.27	−20.9	−21.9		
10	3,4-Dimethylpyridine	2.24	−19.1	−19.7		
11	3,5-Dimethylpyridine	2.21	−18.9	−19.5		
12	3-Aminopyridine	2.20	−18.6	−19.2		
13	2.6-Dimethylpyridine	2.14	−19.5	−20.0	−20.3	−28.5
14	4-Methoxypyridine	2.13	−18.8	−19.3		
15	2-Aminopyridine	2.12	−17.9	−18.4		
16	4-*tert*-Butylpyridine	2.11	−19.1	−19.7		
17	2-Methylaminopyridine	2.11	−18.0	−18.5		
18	4-Methylpyridine	2.07	−18.9	−19.5		
19	4-Ethylpyridine	2.07	−18.9	−19.5		
20	2-Methylpyridine	2.03	−19.0	−19.5		
21	3-Ethylpyridine	2.01	−18.7	−19.3		
22	3-Methylpyridine	2.00	−18.6	−19.2		
23	4-Phenylpyridine	1.96	−18.4	−18.9		
24	Acridine	1.95	−18.0	−18.7		
25	4-Vinylpyridine	1.95	−18.5	−19.0		
26	Isoquinoline	1.94	−18.2	−18.8		
27	Pyridazine	1.95	−19.5	−19.9		
28	2-Ethylpyridine	1.94	−18.9	−19.4		
29	Quinoline	1.89	−18.0	−18.6		
30	2-Butylpyridine	1.88	−18.9	−19.4		
31	Phenanthridine	1.87	−17.7	−18.3		
32	1,7-Phenanthroline	1.87	−17.4	−17.9		
33	Pyridine	1.86	−18.4	−18.9	−20.7	−27.5
34	2-Aminopyrimidine	1.85	−16.5	−16.7		
35	2-Isopropylpyridine	1.76	−18.7	−19.2		
36	2-Vinylpyridine	1.65	−16.9	−17.3	−16.9	−25.0
37	2-N,N-Dimethylaminopyridine	1.61	−17.9	−18.4		
38	Quinazoline	1.55	−16.3	−16.8		
39	4-Chloropyridine	1.54	−16.6	−16.9		
40	Phenazine	1.52	−15.3	−15.7		
41	4-Acetylpyridine	1.41	−16.0	−16.3		
42	Methyl nicotinate	1.45	−16.0	−16.4		
43	3-Benzoylpyridine	1.41	−16.9	−17.4		
44	2,2'-Bipyridine	1.45	−17.7	−18.2		
45	2-Phenylpyridine	1.43	−19.0	−19.6	−18.8	−27.8

Table 1. Continued.

No.	Compound	pK_β	MIP_{AM1}	MIP_{CCl4}	$MIP_{HF/6-31G*}$	MIPp
46	2-*tert*-Butylpyridine	1.42	−18.3	−18.8	−18.2	−27.4
47	Pyrimidine	1.37	−15.9	−16.3		
48	3-Fluoropyridine	1.35	−16.2	−16.5		
49	3-Chloropyridine	1.31	−16.5	−16.9		
50	Pyrazine	1.22	−15.2	−15.3		
51	7,8-Benzoquinoline	1.16	−15.0	−15.5	−15.1	−24.3
52	2-Chloropyridine	1.05	−18.6[b]	−19.2	−19.1	−26.1
53	4-Cyanopyridine	0.92	−14.7	−14.8		
54	3-Cyanopyridine	0.82	−14.7	−14.8		
55	2-Methoxypyridine	0.99	−15.7	−16.0		
56	2-Fluoropyridine	0.95	−18.1[b]	−18.8	−19.4	−25.6
57	2-Cyanopyridine	0.48	−16.1[b]	−16.6	−16.0	−23.1
58	3,5-Dichloropyridine	0.85	−14.7	−15.0		
59	*s*-Triazine	0.80	−13.4	−13.5	−14.4	−20.5
60	2,6-Difluoropyridine	0.14	−17.9[b]	−18.6	−17.6	−23.8
61	Pentafluoropyridine	∼ −0.49	−11.9[b]	−12.1	−11.4	−17.4

[a]Data taken from ref. 68.
[b]The MIP values obtained after correcting the semiempirical AM1 electrostatic potential by a factor of 0.82 for the *ortho*-substituted halogen/cyano derivatives are −14.8 (**52**), −14.4 (**56**), −12.5 (**57**), −11.1 (**60**) and −8.8 (**61**) (see text for details).

values are close, but slightly more negative than the values computed using *in vacuo* wavefunctions. The pK_β *versus* MIP_{CCl4} regression equation ($r = 0.92$) shows, however, no remarkable statistical improvement compared to Equation 9.

The effect of computing the MIP at the *ab initio* HF/6-31G(d) level was explored for a subset of compounds covering the whole range of pK_β: pyridine and its 4-amino, 2,6-dimethyl, 2-vinyl, 2-phenyl and 2-*tert*-butyl derivatives, *s*-triazine, and 7,8-benzoquinoline, as well as 2-fluoro, 2-chloro, 2-cyano, 2,6-difluoro and pentafluoropyridine. For this subset of compounds the use of the HF/6-31G(d) wavefunction improves the ability of the MIP to reproduce changes in pK_β, as noted in the increase in Pearson's correlation coefficient (see Equations 10 and 11). Finally, inclusion of the polarization contribution computed at the HF/6-31G(d) (see Equation 2) further increases the percentage of variance in the pK_β values for the subset of compounds (Equation 12).

$$pK_\beta = -0.25(0.07)MIP_{AM1} - 3.06(1.21)$$
$$r = 0.73 \quad r^2_{cv} = 0.35 \quad n = 13 \quad F = 12.5 \qquad (10)$$

$$pK_\beta = -0.22(0.05)MIP_{HF/6-31G(d)} - 2.73(0.86)$$
$$r = 0.81 \quad r^2_{cv} = 0.55 \quad n = 13 \quad F = 21.1 \qquad (11)$$

$$pK_\beta = -0.21(0.03)MIPp - 4.13(0.87)$$
$$r = 0.88 \quad r^2_{cv} = 0.70 \quad n = 13 \quad F = 38.1 \qquad (12)$$

The preceding results reveal that MIPp calculations performed at the *ab initio* HF/6-31G(d) level largely correct the deviation of the *ortho*-substituted halogen/cyano derivatives found at the semiempirical level. Nevertheless, the use of HF/6-31G(d) wavefunctions and the inclusion of the polarization energy term make MIPp calculations very expensive, which drastically limits their suitability to predict the hydrogen-bonding properties of large molecules or to screen a large number of compounds. For our purposes, however, the discrepancies found at the semiempirical level might be empirically corrected by adjusting the van der Waals parameters or by scaling the electrostatic energy term. Based on our previous studies [64], this latter strategy was finally adopted and such a deviation was corrected by applying a scaling factor of 0.82 per each halogen/cyano group in *ortho* position. This strategy allows us to predict with reasonable accuracy the pK_β values of the N-heterocycles, including the *ortho*-substituted halogen/cyano derivatives, at a low computational cost, as can stated from Equation 13 (see Figure 1).

$$pK_\beta = -0.25(0.01)MIP_{AM1} - 2.62(0.21)$$
$$r = 0.95 \quad r^2_{cv} = 0.90 \quad n = 61 \quad F = 598.6 \qquad (13)$$

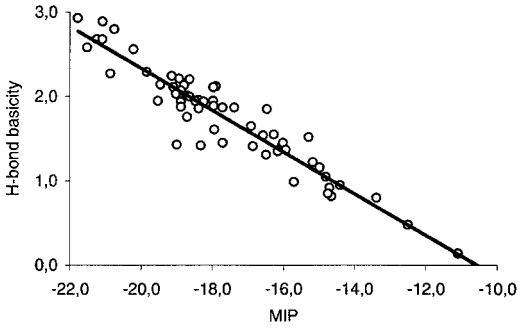

Figure 1. Representation of the pK_β values in front of the MIP/AM1 (kcal/mol) energies for a series of N-heteroaromatic compounds [taken from ref. 68].

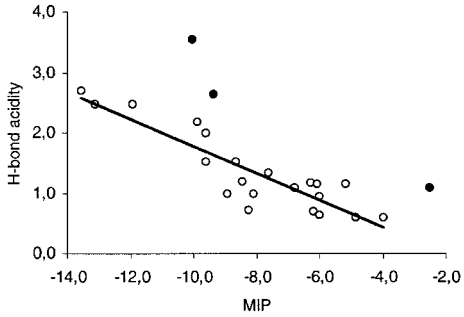

Figure 2. Representation of the pK_α values in front of the MIP/AM1 (kcal/mol) energies for a series of NH hydrogen-bond donor compounds [taken from ref. 69]. Compounds excluded from the regression equation are shown in black.

In summary, the preceding results allows us to conclude that the MIP functional determined from the AM1 semiempirical wavefunction is useful to predict the hydrogen-bond strength for the series of N-heteroaromatic derivatives.

pK_α of NH and OH hydrogen-bond donors

Table 2 reports the pK_α values of 24 NH and 32 OH hydrogen-bond donors reported by Abraham et al. [69]. The NH series, which covers a range of pK_α values varying from 0.60 to 3.55, includes diverse functional groups: amines, amides and tioamides, sulphonamides, and heterocycles, like pyrrole, indole, and substituted derivatives of imidazole, triazole and tetrazole.

The variation of the pK_α values *versus* the MIP/AM1 energies *in vacuo* is shown in Figure 2. Removing compounds **13**, **14** and **24** (see NH donors in Table 2), there is a clear linear relationship between the pK_α and MIP values (see Figure 2 and Equation 14). The deviation of compounds **13**, **14** and **24** is intriguing, though it might be related to conformational effects, as in the case of the phenyl group attached to the tetrazole ring of compound **24**, or to the occurrence of various tautomeric forms.

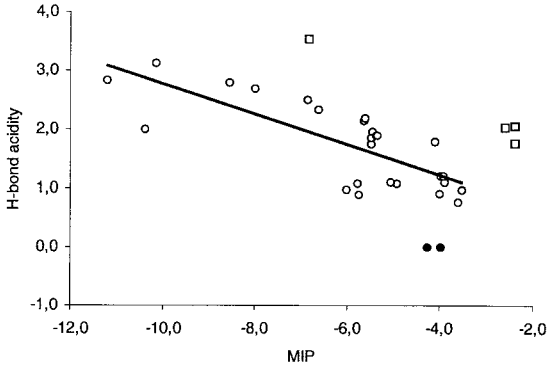

Figure 3. Representation of the pK_α values in front of the MIP/AM1 (kcal/mol) energies for the series of OH hydrogen-bond donor (aliphatic alcohol, oximes and phenols: circle; carboxilic acids: square) compounds [taken from ref. 69]. Compounds excluded from the regression equation are shown in black.

$$pK_\alpha = -0.22(0.03)MIP_{AM1} - 0.45(0.22)$$
$$r = 0.89 \quad r^2_{cv} = 0.76 \quad n = 21 \quad F = 73.4 \qquad (14)$$

The series of 32 OH hydrogen-bond donors (see Table 2), which spans a range of pK_α values comprised between 0.0 and 3.55, includes alcohols, oximes, phenols, and carboxylic acids. When considering the whole data set, poor correlations between pK_α and MIP values were found. However, a reasonable regression model can be obtained after exclusion of carboxylic acids, as noted in Equation 15 (see also Figure 3), which gives the regression line obtained for a subset of 26 compounds including alcohols, oximes and phenols. Inspection of Figure 3 suggests that carboxylic acids seem to be systematically deviated from the rest of compounds. It can be speculated that this effect likely reflects the different chemical environment of the OH group in carboxylic acid compounds, where the lone pairs on the proximal carbonyl oxygen might affect the approach of the hydrogen-bond acceptor compound (N-methylpyrrolidine in ref. 69). However, caution is necessary owing to the limited number of carboxylic acid compounds and to their nonuniform distribution.

$$pK_\alpha = -0.26(0.06)MIP_{AM1} + 0.21(0.22)$$
$$r = 0.77 \quad r^2_{cv} = 0.52 \quad n = 26 \quad F = 22.4 \qquad (15)$$

In order to derive Equation 15, compounds 2,6-diisopropylphenol (**39**) and 2,6-di-*tert*-butylphenol (**41**) were not considered because they clearly deviate from the regression line (see Figure 3). Such a

Table 2. Hydrogen-bond acidity[a] (pK_α) of NH and OH donor compounds and the interaction energy (kcal/mol) determined from MIP calculations by using the AM1 wavefunction *in vacuo* (MIP). in chloroform (MIP_{CHCL3}) or in octanol ($MIP_{octanol}$) and from the HF/6-31G(d) wavefunction without ($MIP_{HF/6-31G(d)}$) or with (MIPp) polarization energy.

No	Name	pK_α	MIP_{AM1}	MIP_{CHCL3}	$MIP_{octanol}$	$MIP_{HF/6-31G*}$	MIPp
	NH donors						
1	1,2,3,4-Tetrahydroquinoline	0.60	−4.0	−4.4	−4.5	−4.4	−6.0
2	10,11-Dihydro-5H-dibenzo[b,f]azepine	0.60	−4.9	−5.0	−5.1	−6.3	−8.8
3	4-Nitro-N-methylaniline	0.73	−8.3	−9.6	−10.2	−10.0	−11.3
4	5-Nitroindoline	1.00	−8.1	−9.5	−10.1	−8.5	−10.8
5	2-Aminobenzothiazole	1.10	−6.8	−7.3	−7.7	−7.4	−8.7
6	CF_3CONH_2	1.52	−9.6	−10.8	−11.4	−10.2	−11.3
7	$C_6H_{13}NHCOC_6H_{13}$	0.64	−6.0	−6.8	−8.2	−7.3	−9.2
8	$MeNHCOBu^t$	0.70	−6.2	−7.1	−7.6	−7.8	−9.0
9	Acetanilide	1.34	−7.6	−8.7	−9.2	−9.7	−11.4
10	3′-Chloro-4′-nitroacetanilide	2.48	−11.9	−13.9	−14.8	−14.4	−16.3
11	3′-Trifluoromethyl-4′-nitroacetanilide	2.47	−13.1	−15.3	−16.2	−15.8	−18.0
12	Thioacetanilide	1.52	−8.6	−10.1	−10.9	−10.0	−12.1
13	1-Oxa-2-azaspiro[4.5]decan-3-one	1.10	−2.5	−2.6	−2.7	−3.7	−4.6
14	$(CF_3CO)_2NH$	2.63	−9.4	−10.5	−10.7	−10.9	−12.5
15	Toluene-p-sulphonamide	1.15	−5.2	−5.6	−5.9	−5.7	−7.1
16	N-(2-Naphthyl)toluene-*p*-sulphonamide	1.18	−6.3	−6.8	−7.0	−8.2	−10.2
17	$C_7H_{15}CONHSO_2Me$	1.00	−8.9	−9.7	−10.0	−7.2	−8.8
18	Pyrrole	0.95	−6.0	−6.7	−7.1	−7.0	−8.0
19	Indole	1.15	−6.1	−6.8	−7.2	−7.4	−8.7
20	Phenyl 4-(1H-imidazol-2yl)butanoate	1.20	−8.4	−9.6	−10.2	−9.3	−10.6
21	3-(3-Phenylpropyl)-4H-1,2,4-triazole	1.99	−9.6	−11.0	−11.7	−11.5	−12.2
22	4-(Methylthio)-1H-1,2,3-triazole	2.18	−9.9	−11.4	−12.2	−12.3	−14.1
23	4-(Trifluoromethyl)-1H-1,2,3-triazole	2.71	−13.6	−15.3	−16.2	−14.4	−16.1
24	Phenyl-1H-tetrazole	3.55	−10.0	−11.3	−12.0	−12.2	−14.3
	OH donors						
25	Ethanol	1.21	−4.0	−4.6	−4.9		
26	Propan-1-ol	1.11	−3.9	−4.4	−4.7		
27	Hexan-1-ol	1.20	−3.9	−4.5	−4.8		
28	Propan-2-ol	0.91	−4.0	−4.4	−4.5		
29	*tert*-Butyl alcohol	0.78	−3.9	−4.2	−4.3		
30	$PhCH_2OH$	0.90	−5.7	−6.2	−6.4		
31	$ClCH_2CH_2OH$	1.08	−5.8	−6.4	−6.7		
32	CF_3CH_2OH	2.00	−10.4	−11.2	−11.7		
33	$(CF_3)_2CHOH$	2.83	−11.2	−12.1	−12.5		
34	Phenol	2.14	−5.6	−6.1	−6.4		
35	2-Metylphenol	1.75	−5.5	−6.0	−6.2		
36	2.6-Dimethylphenol	1.08	−4.9	−5.3	−5.5		
37	2-Isopropylphenol	1.95	−5.5	−5.9	−6.1		
38	2.6-Diisopropylphenol	0.00	−4.3	−4.5	−4.5		
39	2-*tert*-Butylphenol	1.85	−5.5	−5.9	−6.1		
40	2.6-Di-*tert*-Butylphenol	0.00	−4.0	−6.2	−7.5		
41	2-Chlorophenol	2.33	−6.6	−7.3	−7.7		
42	2.6-Dichlorophenol	0.98	−3.5	−3.5	−3.6		

[a]Data taken from ref. 69.

Table 2. Continued

No	Name	pK_α	MIP_{AM1}	MIP_{CHCL3}	$MIP_{octanol}$	$MIP_{HF/6-31G*}$	MIPp
43	2-Cyanophenol	2.69	−8.0	−9.0	−9.4		
44	3-N,N-Dimethylaminophenol	1.79	−4.1	−4.3	−4.5		
45	3-Methylphenol	1.89	−5.4	−5.8	−6.1		
46	3-Isopropylphenol	1.89	−5.4	−5.8	−6.0		
47	3-Chlorophenol	2.50	−6.9	−7.7	−8.1		
48	4-Methoxyphenol	2.18	−5.6	−6.3	−6.6		
49	4-Trifluoromethylphenol	2.80	−8.6	−9.6	−10.1		
50	4-Nitrophenol	3.12	−10.2	−11.6	−12.2		
51	(2E)-4a-Phenyloctahydronapthalen-2(1H)-one oxime	0.98	−6.0	−6.4	−6.7		
52	(1E)-2,2-Dimethyl-1-phenylheptan-1-one oxime	1.11	−5.1	−5.3	−5.5		
53	Acetic acid	2.04	−2.6	−2.8	−3.0		
54	Pivalic acid	1.77	−2.4	−2.6	−2.7		
55	Benzoic acid	2.07	−2.4	−2.5	−2.6		
56	Trifluoroacetic acid	3.55	−6.8	−7.5	−7.9		

deviation can be related to the fact that replacing N-methylpyrrolidine (i.e., the reference compound used in the experimental determination of the hydrogen-bond acidity scale [69]) by a single point particle in the MIP calculation is a crude approximation to describe properly the steric effects due to the presence of bulky substituents in 2,6-disubstituted phenols. This finding, therefore, suggests that caution is necessary regarding the use of experimental values, since differences in experimentally determined hydrogen-bonding scales might originate from the use of bigger or smaller hydrogen-bond acceptors/donors as reference compounds.

It is worth noting that Equation 15 largely reflects the dependence between the hydrogen-bond acidity and the MIP functional for the phenol derivatives, as can be stated in Equation 16. The greater predictive power of the MIP for phenols relative to alcohols and oximes can be explained by two factors. First, the largest chemical diversity spanned by the subset of phenol compounds, which is reflected in the variation of the hydrogen-bond acidity values from 0.98 (**42**) to 3.12 (**50**) while the subset of alcohols/oximes are chemically less diverse (with exception of compounds 32 and 33, the hydrogen-bond acidity of the rest of compounds lies in the range 0.90–1.21). Second, the greater conformational flexibility of the alcohols in solution, whereas theoretical calculations were limited to a single species whose aliphatic chain was always built up in an *all-trans* conformation.

$$pK_\alpha = -0.31(0.02)MIP_{AM1} + 0.19(0.15)$$

$$r = 0.91 \quad r^2_{cv} = 0.78 \quad n = 15 \quad F = 103.9 \quad (16)$$

The potential influence of solvation was examined from MIP calculations performed by using the MST/AM1 wavefunctions determined in chloroform ($\varepsilon = 4.71$) and octanol ($\varepsilon = 9.87$). However, the regression equations for either NH- or OH-donor compounds did not improve the percentage of variance in the pK_α values explained by the linear regression models obtained from MIP/AM1 calculations *in vacuo*. Similarly, the influence of the MIP or MIPp calculated at the HF/6-31G(d) level was explored for the NH series of compounds. However, in contrast with the behavior found for the N-heteroaromatic compounds (see above), no improvement in the quality of the pK_α *versus* MIP/AM1 model was found.

pK_β of N and O hydrogen-bond acceptors

Table 3 reports the pK_β values of 47 N and 36 O hydrogen-bond acceptors taken from the data compiled by Abraham et al. [69]. The pK_β values for the N series range from 0.61 to 3.68 and correspond to a broad variety of functional groups, like amines, nitriles, and diverse N-containing heteroaromatics. Table 3 gives the pK_β values of these compounds together with the MIP energies.

The variation of the pK_β values with the MIP/AM1 energies is represented in Figure 4, which shows the existence of a linear correlation between the two sets of data, as noted in Equation 17. Following our previous findings for *ortho*-substituted halogen/cyano

Table 3. Hydrogen-bond basicity[a] (pK_β) of N and O acceptor compounds and the interaction energy (kcal/mol) determined from MIP calculations by using the AM1 wavefunction *in vacuo* (MIP).

No	Compound	pK_β	MIP
	N acceptors		
1	Isopropylamine	2.84	−20.9
2	Benzylamine	2.36	−18.9
3	Allylamine	2.63	−19.7
4	$CN(CH_2)_2NH_2$	1.74	−16.7
5	$CF_3CH_2NH_2$	1.01	−12.2
6	Pyridine	2.52	−18.4
7	2-Methoxypyridine	1.28	−15.7
8	2-Fluoropyridine	1.41	−18.1[b]
9	2-Chloropyridine	1.48	−18.6[b]
10	2-Cyanopyridine	1.00	−16.1[b]
11	3-Methylpyridine	2.65	−18.6
12	3-Fluoropyridine	1.82	−16.1
13	3-Chloropyridine	1.77	−16.5
14	3-Cyanopyridine	1.41	−14.7
15	4-Methylpyridine	2.78	−18.9
16	3,4-Dimethylpyridine	3.06	−19.1
17	4-Methoxypyridine	2.87	−18.8
18	4-N.N-Dimethylaminopyridine	3.54	−20.8
19	Pyrazine	1.46	−15.2
20	Pyrimidine	1.67	−15.9
21	Pyridazine	2.53	−19.5
22	Isoxazole	1.06	−15.4
23	Oxazole	1.67	−17.6
24	2.4.5-Trimethyloxazole	2.65	−19.3
25	Thiazole	1.90	−18.1
26	Benzothiazole	1.76	−17.0
27	1-Methylpyrazole	2.22	−16.7
28	1-Methylimidazole	3.68	−22.0
29	1-Benzyl-1.2.4-triazole	2.38	−18.7
30	1-Phenethyl-1.2.3-triazole	2.56	−19.8
31	1-Methylbenzotriazole	2.17	−18.0
32	4-Butyl-4H-1,2,4-triazole	3.37	−22.6
33	3,5-Diphenyl−1,2,4-oxadiazole	0.57	−13.5
34	2,5-Dimethyl-1,3,4-thiadiazole	2.51	−20.0
35	2-(Methylthio)-5-Methyl-1,3,4-thiadiazole	1.98	−18.6
36	2,1,3-Benzothiadiazole	0.79	−14.6
37	N-cyclohex-3-en-1-yl-O-nonylhydroxylamine	1.51	−14.6
38	$Me_2C{=}NOPh$	1.10	−12.9
39	3,3-Bis(dimethylamino)acrylonitrile	2.90	−19.0
40	Me_2NCN	2.00	−17.6
41	Acetonitrile	1.23	−15.9
42	$MeOCH_2CN$	1.04	−15.2
43	$MeO(CH_2)_2CN$	1.28	−15.1
44	$ClCH_2CN$	0.61	−13.2
45	PhCN	1.06	−15.4

Table 3. Continued.

No	Compound	pK_β	MIP
46	4-Methoxybenzonitrile	1.32	−15.9
47	4-Chlorobenzonitrile	0.92	−14.3
	O acceptors		
48	Ethanol	1.41	−13.8
49	Propan-2-ol	1.36	−14.2
50	tert-Butyl alcohol	1.45	−14.3
51	Dibutyl ether	1.28	−13.8
52	tert-Butyl methyl ether	1.46	−13.4
53	Tetrahydrofuran	1.69	−14.4
54	Anisole	0.30	−10.0
55	$MeO(CH_2)_2OMe$	1.69	−12.1
56	1-4-Dioxane	1.28	−12.4
57	1.3-Dioxolane	0.70	−11.2
58	Acetone	1.61	−15.8
59	Pentan-3-one	1.50	−15.7
60	$MeCOPr^i$	1.52	−15.6
61	$MeCOBu^t$	1.44	−15.8
62	Pr^iCOPr^i	1.39	−15.6
63	Cyclohexanone	1.70	−15.9
64	Acetophenone	1.46	−15.4
65	Ethyl acetate	1.43	−15.9
66	Dihydro-2(3H)-thiophenone	1.32	−14.4
67	Dimethylformamide	2.81	−18.5
68	Diethylformamide	2.73	−18.4
69	$Bu^tCON(Me)Bu^t$	2.53	−18.5
70	N-Methylpyrrolidinone	3.12	−18.7
71	N-Dimethylbenzamide	2.82	−18.6
72	Tetramethylurea	3.19	−17.9
73	2-Methylisoxazolidin-3-one	2.38	−15.7
74	$PhOCONMe_2$	2.09	−17.2
75	N-Methylmaleimide	1.67	−12.5
76	Dimethyl sulphoxide	3.06	−25.9
77	Tetramethylenesulphone	1.61	−21.5
78	$PhSO_2N(Me)CH_2Ph$	1.36	−23.1
79	2-Benzyl-1-benzothiophen-3(2H)-one 1,1-dioxide	0.99	−21.0
80	Triphenylphospine oxide	3.85	−37.5
81	Triethylphosphate	3.17	−35.6
82	3-N.N-Diethylcarbamoylpyridine	2.76	−17.7
83	4-Acetylpyridine	2.20	−16.0

[a]Data taken from ref. 69.
[b]The MIP values obtained after correcting the semiempirical AM1 electrostatic potential by a factor of 0.82 for the *ortho*-substituted halogen/cyano derivatives are −14.0 (**8**), −14.3 (**9**) and −12.3 (**10**) (see text for details).

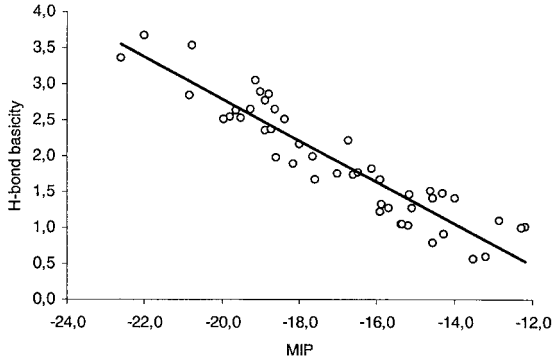

Figure 4. Representation of the pK_β values in front of the MIP/AM1 (kcal/mol) energies for the series of N hydrogen-bond acceptor compounds [taken from ref. 69].

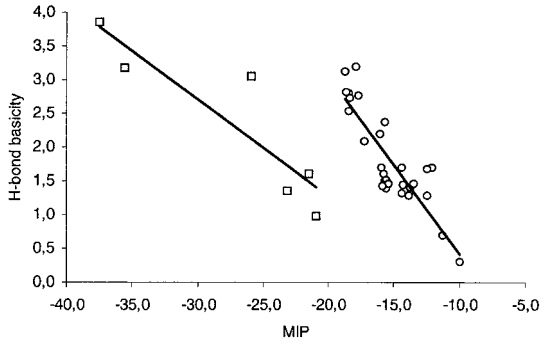

Figure 5. Representation of the pK_β values in front of the MIP/AM1 (kcal/mol) energies for the series of O hydrogen-bond acceptor (alcohol, ether and carbonyl: circle; P=O and S=O: square) compounds [taken from ref. 69].

derivatives of pyridine (see above), the electrostatic component of the semiempirical AM1 MIP functional for compounds 2-fluoro, 2-chloro and 2-cyanopyridine was scaled by a factor of 0.82, which improved the fitting of these compounds into the regression model.

$$pK_\beta = -0.29(0.02)MIP_{AM1} - 2.99(0.30)$$

$$r = 0.93 \quad r^2_{cv} = 0.85 \quad n = 47 \quad F = 278.5 \quad (17)$$

Comparison of Equation 17 with the regression model obtained for the N-heteroaromatics compounds examined by Berthelot et al. (Equation 13) reveals a qualitative agreement, though the slope in Equation 17 (slope: −0.29) is slightly larger than in Equation 13 (slope: −0.25). This difference can be attributed to the different polarity of the two solvents. This is noted in the regression equation (Equation 18) obtained for a subset of 16 heteroaromatic compounds included in the two series of N-acceptor compounds, which indicates an enhancement of the pK_β values in 1,1,1-trichloroethane compared to the values determined in carbon tetrachloride.

$$pK_\beta(CH_3CCl_3) = 1.19(0.05)pK_\beta(CCl_4) + 0.24(0.09)$$

$$r = 0.99 \quad r^2_{cv} = 0.96 \quad n = 16 \quad F = 506.8 \quad (18)$$

The series of 36 O-acceptor compounds (see Table 3) spans a range of pK_β values comprised between 0.3 and 3.85. It includes alcohols, ethers, and diverse X=O bases (X=C, P, S). Compounds 3-N,N-diethylcarbamoylpyridine (**82**) and 4-acetylpyridine (**83**) were also included, as there is evidence showing that the major site for hydrogen bonding is the carbonyl oxygen [69]. The series of compounds bearing P=O and S=O groups appeared clearly differentiated from the rest of compounds (see Figure 5), which likely stems from deficiencies of the AM1 hamiltonian to reflect the balance between X=O and X^+-O^- forms for compounds with X=C relative to compounds with X=P or S [70]. Accordingly, linear regression models were determined separately for compounds bearing i) alcohol, ether and carbonyl (Equation 19) and ii) P=O and S=O (Equation 20) groups. However, the regression model given by Equation 20 has to be treated with caution owing to the reduced number of compounds included in this latter subset of hydrogen-bond acceptors.

Alcohol, ether and carbonyl:

$$PK_\beta = -0.26(0.03)MIP_{AM1} - 2.22(0.45)$$

$$r = 0.86 \quad r^2_{cv} = 0.71 \quad n = 30 \quad F = 82.5 \quad (19)$$

P=O and S=O:

$$pK_\beta = -0.14(0.04)MIP_{AM1} - 1.60(0.98)$$

$$r = 0.90 \quad r^2_{cv} = 0.66 \quad n = 6 \quad F = 16.6 \quad (20)$$

The preceding results shows that the MIP/AM1 method is useful to describe the hydrogen bonding characteristics of N- and O-acceptor compounds. Based on the results reported before for the series of N-heteroaromatic compounds, little improvement is expected upon inclusion of solvent effect. Likewise, though the use of *ab initio* wavefunctions and the inclusion of polarization energy component can improve the predictive ability of hydrogen-bond basicities (see above), their application to large-sized molecules or to large number of compounds will be seriously limited.

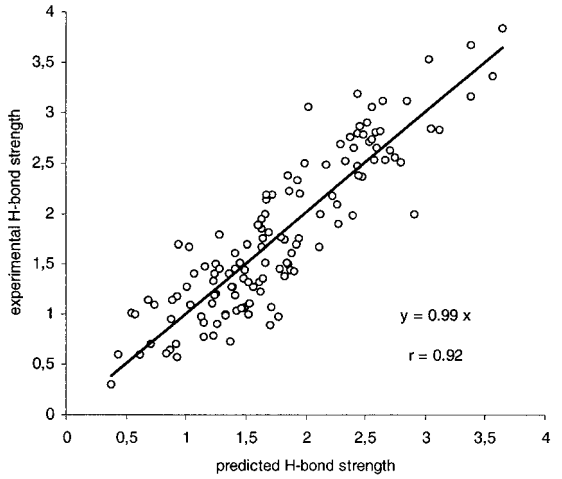

Figure 6. Representation of the experimental versus theoretical pK_α and pK_β values for the 130 hydrogen-bond donors and acceptor compounds taken from ref. 69.

Final remarks

The results presented above highlight the existence of a strong relationship between the propensity for hydrogen-bonding strength and the MIP interaction energy computed at the semiempirical AM1 level. This can be stated from inspection of Figure 6, which represents the experimental pK_α/pK_β values taken from the data compiled by Abraham et al. [69] in front of the theoretical values predicted from the MIP/AM1 energies. The scaling coefficient of the equation $y = cx$ (y and x stand for experimental and predicted values) is close to unity ($r = 0.92$; $r^2_{cv} = 0.85$). The average deviation between experimental and predicted pK_α and pK_β values is 0.01, and the root-mean square deviation amounts to 0.3 units for a total set of 130 compounds.

Comparison of the regression models indicate that the slope of the regression is relatively similar for the hydrogen-bond donor/acceptor compounds. Thus, for the series of compounds whose hydrogen-bond acidity/basicity was examined in 1,1,1-trichloroethane [69], the slope of the regression model amounts to −0.22 for NH-donors (Equation 14), −0.26 for OH-donors (Equation 15; −0.31 for the subset of phenol derivatives in Equation 16), −0.29 for N-acceptors (Equation 17) and −0.26 for O-acceptors (Equation 18, where P=O and S=O compounds were not included). This suggests that for the compounds examined in the study the attached substituents exert a similar influence on the hydrogen-bonding propensity and that such influence is captured in the variation of the MIP energy values. In contrast, there is more variance in the intercept of the regression models, which amounts to −0.45 for NH-donors (Equation 14), +0.21 for OH-donors (Equation 15; +0.19 for the subset of phenol derivatives in Equation 16), −2.99 for N-acceptors (Equation 17) and −2.22 for O-acceptors (Equation 18). Such a variance can be mostly attributed to the intrinsic differences in the chemical nature of the different hydrogen-bond donor/acceptor groups.

The results presented above point out that the MIP functional computed by using the *gas phase* wavefunction of the hydrogen-bond acceptor/donor is useful to predict the variation in experimental hydrogen-bond acidity/basicity determined in 1,1,1-trichloroethane and carbon tetrachloride. This finding, which was expected in carbon tetrachloride, is more remarkable for the data measured in 1,1,1-trichloroethane, a solvent chosen as a better model for real biological environments than less polar solvents [69]. It is worth noting, nevertheless, that the slope in the regression model derived for the N-heteroaromatic hydrogen-bond acceptors in 1,1,1-trichloroethane (Equation 17) is larger than that found in carbon tetrachloride (Equation 13), and that the ratio between the slopes agrees with the ratio observed between the hydrogen-bond scales determined in the two solvents (Equation 18). Based on these results, it can be expected that the computational procedure presented here could be useful in drug design studies, as it can be exploited to design chemical modifications that modulate the intrinsic hydrogen-bond donor/acceptor capabilities that mediate the interaction of drugs in the binding site of biological receptors [71].

From an experimental point of view, the determination of the hydrogen-bond propensities in polar solvents like water is handicapped by the unfavorable role of solvation in high dielectric media, which destabilizes hydrogen-bonded complexes and can even alter the specific nature of the intermolecular association [72, 73]. Therefore, hydrogen-bonding capabilities must in principle be extrapolated from the data collected in organic solvents. In contrast, the computational procedure described here should allow the exploration of the 'intrinsic' hydrogen-bonding capabilities of compounds in water, since the inclusion of hydration effects is accomplished through the use of the wavefunction determined from MST calculations in water [74, 75]. Since hydrogen-bonding has been described to be important in modulating properties like drug solubility [76, 77], absorption and permeation [78–80], it can be expected that a more precise de-

scription of the hydrogen-bond formation propensities in water will be valuable to further examine the role of hydrogen bonding in biological and pharmacological properties of bioactive compounds.

To this end, it is worth noting that the application of the MIP to molecules containing diverse functional groups is straigthforward. At this point, let us recall that the calculation of the MIP requires only knowledge of the wavefunction of the isolated hydrogen-bond donor/acceptor molecule, thus avoiding any information about the hydrogen-bonded complex. The mutual electronic influence of the various substituents attached to the molecule is naturally included in the molecular wavefunction (see Equation 1), and the steric hindrance between proximal groups is accounted for by means of the van der Waals term (see Equation 4). Therefore, a global measure of the molecular hydrogen-bond donor/acceptor potential can be derived from the MIP minima obtained for the different functional groups present in the molecule. Finally, calculation of the MIP in grid of points surrounding the molecule can also be useful to visualize the 3D distribution of hydrogen bonding capabilities of molecules. This kind of calculations, which imply the evaluation of the electrostatic potential over a large number of points, might benefit from approaches based on a dual quantum mechanical/molecular mechanical description of the molecule [81].

Based on the preceding discussion, we believe that present results can be considered as a starting point for further exploration of the MIP as a tool to examine the role of hydrogen-bonding in drug design and structure-activity relationships. Since the prediction of hydrogen-bonding propensities can be reasonably achieved even at the semiempirical level, the method should be applicable to medium-sized molecules. The benefits of using the MIP procedure in these areas is currently being investigated.

Acknowledgements

We are grateful to Prof. J. Tomasi for providing us with his original code of the PCM model, which was modified by us to carry out the MST calculations. We also acknowledge the Dirección General de Investigación Científica y Técnica (grants PB98-1222 and PM99-0046) for financial support, and the Centre de Supercomputació de Catalunya for computational facilities.

References

1. S. Scheiner. Molecular Interactions. From van der Waals to Strongly Bound Complexes. Wiley. Chichester 1997.
2. G. A. Jeffrey. An introduction to Hydrogen Bonding. Oxford University Press. New York 1997.
3. G. R. Desiraju, T. Steiner. The Weak Hydrogen Bond. Oxford University Press. Oxford 1999.
4. S. Scheiner. Hydrogen Bonding. Oxford University Press. New York 1997.
5. J. M. Lehn. Supramolecular Chemistry: Concepts and Perspectives. VCH. Weinheim. 1995.
6. G. A. Jeffrey, W. Saenger. Hydrogen Bonding in Biological Structures. Springer. Berlin 1991.
7. I. Alkorta, I. Rozas, J. Elguero. Chem. Soc. Rev., 27 (1998) 163
8. P. Hobza, Z. Havlas. Chem. Rev., 100 (2000) 4253
9. E. Cubero, M. Orozco, P. Hobza, F. J. Luque,. J. Phys. Chem. A., 103 (1999) 6394.
10. E. Cubero, M. Orozco, F. J. Luque. Chem. Phys. Lett., 310 (1999) 445.
11. W. L. Jorgensen. Chemtracts-Org Chem., 4 (1991) 91.
12. P. A. Grey,. S. A. Whitt, J. B. Tobin,. Science, 264 (1994) 1927.
13. J. A. Gerlt, P. G. Gassman.. J. Am. Chem. Soc., 115 (1993) 11552.
14. W. W. Cleland, M. M. Kreevoy. Science, 264 (1994) 1887.
15. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney. Adv. Drug Deliv. Rev. 23 (1997) 3.
16. 3D QSAR in Drug Design: Theory, Methods and Applications. H. Kubinyi (ed.). ESCOM. Leiden. 1993.
17. P. J. Gane, P. M. Dean. Curr. Op. Struct. Biol. 10 (2000) 401.
18. W. Saenger. Principles of Nucleic Acid Structure. Springer-Verlag. New York, NY. 1984
19. J. Sadowski, K. Kubinyi, K. J. Med. Chem. 41 (1998) 3325.
20. Ajay, W. P. Walters, M. A. Murcko, M.A. J. Med. Chem. 41 (1998) 3314.
21. T. Fujita, T. Nishioka, M. Nakajima. J. Med. Chem. 20 (1977) 1071.
22. M. Charton, B. I. Charton. J. Theor. Biol. 99 (1982) 629.
23. M. Rueda, F. J. Luque, M. Orozco. Biopolymers, in press.
24. M. J. Kamlet, R. M. Doherty, J. L: M. Abboud, M. H. Abraham, R. W. Taft. J. Pharm. Sci. 75 (1986) 338.
25. M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris, P. J. Taylor. J. Chem. Soc., Perkin Trans 2 (1989) 699.
26. M. H. Abraham. Chem. Soc. Rev. 22 (1993) 73
27. P. Seiler. Eur. J. Med. Chem. 9 (1974) 473.
28. R. C. Young, R. C. Mitchell, T. H. Brown, C. R. Ganellin, R. Griffiths, M. Jones, K. K. Rana, D. Saunders, I. R. Smith, N. E. Sore, T. J. Wilks. J. Med. Chem. 31 (1988) 656.
29. I. Moriguchi. Chem. Pharm. Bull. 23 (1975) 247.
30. L. Y. Wilson, G. R. Famini. J. Med. Chem. 34 (1991) 1668.
31. J. C. Dearden, M. T. D. Cronin, D. Wee. J. Pharm. Pharmacol. 49, Suppl. 4 (1997) 110.
32. J. S. Murray, P. Politzer. J. Org. Chem. 56 (1991) 6715.
33. E. Gancia, J. G. Montana, D. T. Manallack. J. Mol. Graphics Mod. 19 (2001) 349.
34. J. A. Platts. Phys. Chem. Chem. Phys. 2 (2000) 973
35. J. A. Platts. Phys. Chem. Chem. Phys. 2 (2000) 3115
36. M. Orozco, F. J. Luque. J. Comput. Chem. 14 (1993) 587.
37. M. Orozco, F. J. Luque. In Molecular Electrostatic Potentials: Concepts and Applications. Theoretical and Computational Chemistry, Vol. 3. J. S. Murray and K. Sen (eds.) Elsevier. Amsterdam. 1996, p. 181

38. F. J. Luque, M. Orozco. J. Comput. Chem. 19 (1998) 866.
39. E. Cubero, F. J. Luque, M. Orozco. Proc. Natl. Acad. Sci. USA 95 (1998) 5976.
40. F. J. Luque, J. M. López, M. L. de la Paz, C. Vicent, M. Orozco. J. Phys. Chem. A 102 (1998) 6690.
41. B. Hernández, F. J. Luque, M. Orozco. J. Comput. Chem. 20 (1999) 937.
42. M. Amat, J. Bosch, J. Hidalgo, M. Cantó, M. Pérez, N. Llor, E. Molins, C. Miravitlles, M. Orozco, F. J. Luque. J. Org. Chem. 65 (2000) 3074.
43. R. Lavilla, O. Coll, J. Bosch, M. Orozco, F. J. Luque. Eur. J. Org. Chem. (2001) 3719.
44. E. Cubero, F. J. Luque, M. Orozco, J. Gao. Submitted
45. R. Bonaccorsi, R. Cimiraglia, P. Palla, J. Tomasi. Int. J. Quantum Chem. 24 (1983) 307.
46. K. Kitaura, K. Morokuma, M. Int. J. Quantum Chem. 10 (1976) 325.
47. See J. Tomasi, B. Mennuci, R. Cammi,. Theoretical and Computational Chemistry, Vol. 3, J. S. Murray and K. Sen (eds.), Elsevier Science, New York, 1996, p. 1.
48. J. Gao. In Reviews in Computational Chemistry. K. B. Lipkowitz, D. B. Boyd (eds.) Vol. 7. VCH. New York. 1996.
49. M. F. Ruiz-Lopez, J. L. Rivail. In Encyclopedia of Computational Chemistry. P.v. R. Schleyer (ed.) Vol. 1. Wiley. New York. 1998. p. 437.
50. E. Scrocco, J. Tomasi. Top. Curr. Chem. 42 (1973) 95.
51. M. M. Francl. J. Phys. Chem. 89 (1985) 428.
52. M. Orozco, M. Bachs, F. J. Luque. J. Comput. Chem. 16 (1995) 563
53. F. J. Luque, Y. Zhang, C. Aleman, M. Bachs, J. Gao, M. Orozco. J. Phys. Chem. 100 (1996) 4269
54. F. J. Luque, C. Aleman, M. Bachs, M. Orozco. J. Comput. Chem. 17 (1996) 806
55. C. Curutchet, M. Orozco, F. J. Luque. J. Comput. Chem. 22 (2001) 1180.
56. S. Miertus, J. Tomasi,. Chem. Phys. 65 (1982) 239
57. R. A. Pierotti. Chem. Rev. 76 (1976) 717.
58. P. Claverie. In Intermolecular Interactions: From Diatomics to Biomolecules. Pullman, B. (ed.) Wiley: Chichester, 1978.
59. S. Miertus, E. Scrocco, J. Tomasi. Chem. Phys. 55 (1981) 117
60. M. J. S. Dewar, E. G. Zoebisch, E. E. Healy, J. J. P. Stewart. J. Am. Chem. Soc. 107 (1985) 3902.
61. P. C. Hariharan, J. A. Pople. Theor. Chim. Acta. 28 (1973) 213.
62. J. J. P. Stewart. MOPAC-6.0. Modified by F. J. Luque, M. Orozco. University of Barcelona (2000).
63. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94. Rev. A.1. Pittsburgh: Gaussian Inc., 1995
64. G. G. Ferenczy, C. A. Reynolds, W. G. Richards. J. Comput. Chem. 11 (1990) 159.
65. C. Alhambra, F. J. Luque, M. Orozco. J. Comput. Chem. 15 (1994) 12.
66. M. Clark, R. D. Cramer III, N. Opdenbosch. J. Comput. Chem. 10 (1989) 982.
67. F. J. Luque, C. Alhambra, M. Orozco. MOPETE computer program. University of Barcelona, Spain. 1999.
68. F. J. Luque, M. Bachs, M. Orozco. J. Comput. Chem. 15 (1994) 847.
69. M. Berthelot, C. Laurence, M. Safar, F. Besseau. J. Chem. Soc., Perkin Trans 2 (1998) 283.
70. M. H. Abraham, P. P. Duce, D. V. Prior, D. G. Barratt, J. J. Morris, P. J. Taylor. J. Chem. Soc., Perkin Trans 2. (1989) 1355.
71. J. J. P. Stewart. J. Comput. Chem. 10 (1989) 209.
72. P. R. Andrews, D. J. Craik, J. L. Martin. J. Med. Chem. 27 (1984) 1648.
73. M. Orozco, F. J. Luque. Chem. Rev. 100 (2000) 4187.
74. C. Colominas, F. J. Luque, M. Orozco. J. Phys. Chem. A 103 (1999) 6200.
75. M. Orozco, M. Bachs, F. J. Luque. J. Comput. Chem. 16 (1995) 563.
76. C. Curutchet, M. Orozco, F. J. Luque. J. Comput. Chem. 22 (2001) 1180.
77. W. L. Jorgensen, E. M. Duffy. Bioorg. Med. Chem. Lett. 10 (2000) 1155
78. E. M. Duffy, W. L. Jorgensen. J. Am. Chem. Soc. 122 (2000) 2878.
79. S. Rey, G. Caron, G. Ermondi, P. Gaillard, A. Pagliara, P. A. Carrupt, B. Testa. J. Mol. Graphics Mod. 19 (2001) 521.
80. O. A. Raevsky, K.J. Schaper, P. Artursson, J. W. McFarland. Quant. Struct-Act. Relat. 20 (2001) 402
81. J. T. Goodwin, R. A. Conradi, N. F. H. Ho, P. S. Burton. J. Med. Chem. 44 (2001) 3721
82. B. Hernandez, F. J. Luque, M. Orozco. J. Comput. Aided Mol. Design 14 (2000) 329.