The Use of Atomic Charges and Orbital Energies as Hydrogen-bonding-donor Parameters for QSAR Studies: Comparison of MNDO, AM1 and PM3 Methods

TARAVAT GHAFOURIAN AND JOHN C. DEARDEN*

*School of Pharmacy, Tabriz Medical Sciences University, Tabriz 51664, Iran and *School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK*

Abstract

Hydrogen-bonding, important in drug–receptor interactions, also determines the solubility and partitioning of drugs between phases. It is, therefore, important to incorporate the effects of hydrogen-bonding in studies of quantitative structure–activity relationships (QSAR).

In this study the atomic charge on the most positively charged hydrogen atom in a molecule and the energy of the lowest unoccupied molecular orbital (LUMO) have been used as a measure of hydrogen-bond-donor capacity. For several hydrogen-bonding acids the Mulliken atomic charges and the energy of the LUMO produced by use of three semi-empirical methods, AM1, PM3 and MNDO, and MNDO electrostatic-potential-derived atomic charges, have been compared in correlations with solvatochromic hydrogen-bonding acidity ($\Sigma \alpha_2^H$).

Atomic charges and LUMO energies, particularly those calculated by use of the AM1 and MNDO methods, were found to correlate well with $\Sigma \alpha_2^H$. They were also found to be good models of hydrogen-bonding in QSAR correlations.

Hydrogen-bonding interactions are important in drug activity, often controlling drug solubility, partitioning and receptor bonding. Because the effects of hydrogen-bonding are difficult to quantify, their inclusion in quantitative structure–activity relationships (QSAR) is often restricted to indicator variables—for example, a score of one is used if the molecule or substituent is capable of hydrogen-bonding, and a score of zero is used if it is not (Fujita et al 1977). The tendency to form hydrogen-bonds can be divided into hydrogen-bond-donor capacity and hydrogen-bond-acceptor capacity. The solvatochromic hydrogen-bonding parameters of Abraham et al (1989, 1990a) have been used for successful correlation of many diverse properties, but the difficulty with these is that they are experimentally derived and their values for complex molecules are not easily obtained, although commercial software based on a group contribution scheme (Platts et al 1999) is currently being developed to calculate them. For QSAR studies it is desirable to have hydrogen-bonding parameters which are readily accessible by the methods of theoretical chemistry.

Because hydrogen-bonding is mainly electrostatic in nature, it seems reasonable that electrostatic interactions could model hydrogen-bonding ability in compounds (Wilson & Famini 1991). The usual (and simplest) representation of the electrostatic properties of a system is through atom-centred point charges. Different quantum mechanical and molecular mechanical methods can be used to calculate the atomic partial charges. The simplest means of accurately representing the molecular charge distribution is to use Mulliken populations from quantum mechanical calculations. A second approach is to use the molecular electrostatic potential (ESP) evaluated at points in space around the molecule as a guide and to fit this to point-charge models (Singh & Kollman 1984). The accuracy of the model depends on how well the ESP is fitted, and on the accuracy of the ESP

Correspondence: J. C. Dearden, School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK.

itself. The accuracy of the calculation of the ESP depends in turn on the quality of the wave-function. It has been shown (Orozco & Luque 1990) that MNDO electrostatic charges reproduce sophisticated ab-initio 6-31G* PD charges (elec-trostatic potential-derived atomic charges) and experimental dipoles very well; use of AM1 char-ges is not so successful.

In a previous study (Dearden & Ghafourian 1993) the CNDO-calculated Mulliken atomic charges were used to predict experimentally mea-sured hydrogen-bond acidities of compounds. The charge parameter (Q_H) was the highest atomic charge on hydrogen atoms in a molecule which, in correlating with α_2^H—the hydrogen-bond acidity parameter (Abraham et al 1989)—showed family-dependent behaviour (Dearden & Ghafourian 1993). As the family-dependent behaviour of basicity- (or acidity-) dependent properties is a result of a different combination of electrostatic and charge-transfer forces involved in any donor-acceptor combination (Maria et al 1987), the energy of the lowest unoccupied molecular orbital (LUMO) of the hydrogen-bond donor (E_{LUMO}) was chosen, in accordance with the procedure of Wilson & Famini (1991), to describe the charge-transfer contribution to hydrogen-bond donor energy (Dearden & Ghafourian 1995). E_{LUMO} gives a measure of the electron affinity of a compound.

In this study three semi-empirical methods implemented in the MOPAC program (Stewart 1990), MNDO, AM1 and PM3, have been used to calculate hydrogen-bond acidity (hydrogen-bond donor capacity) parameters. These methods are more sophisticated than CNDO because they take lone-pair-lone-pair repulsions into account; a properly parametrized MNDO-type model should, therefore, perform better than an equivalent CNDO model. These methods are all parametric approa-ches and their quality depends not only on the theoretical framework but also on the set of para-meters used in them. The parameters calculated by these methods are the MNDO-calculated electro-static potential-derived (PD) atomic charges, Mulliken atomic charges and E_{LUMO}.

Methods

The program Cobra (Oxford Molecular Limited (OML)) was used to perform conformational ana-lysis on the structures for which $\Sigma\alpha_2^H$ values were available. The lowest-energy conformation was further minimized by use of the Cosmic force field in the Pimms program (OML); MOPAC 6.0 (QCPE, Bloomington, IN) was then used to calcu-late parameters. Atomic charges and highest occupied molecular orbital (HOMO) and LUMO energies were calculated by the three semi-empirical methods in the program, namely MNDO, AM1 and PM3. The program Rattler (OML) was employed to calculate PD atomic charges for the energy-minimized structures using the vectors, geometry and dipoles of the molecule provided by a MOPAC 6.0 (MNDO) output file. To calculate PD atomic charges, ESPs were calculated on three layers at 0·2-Å intervals. A scaling factor of 1·6 was applied to the van der Waals radii to create the dimension of the innermost surface. The dot den-sity on these surfaces was 1 dot Å^{-2}, with the total number of points for each molecule being between 250 and 5000. These programs were all run on a Silicon Graphics workstation.

The highest atomic charge on a hydrogen atom (Q_H) in a molecule was selected. Data were then transferred to the Minitab statistical program (run-ning on a VAX computer) where the theoretical parameters were correlated against $\Sigma\alpha_2^H$—the effective or summation solute hydrogen-bond acidity parameter which is suitable for solvation situations (Abraham 1993).

Results and Discussion

The results of these calculations and the corre-sponding $\Sigma\alpha_2^H$ values, are listed in Table 1. The choice of the scaling factor used in Rattler is jus-tified by the finding that the calculated point-charge models are rather insensitive to which shell(s) is (are) used (Singh & Kollman 1984). PD charges have also been shown to be unaffected by the total number of points, especially when the fitting is performed with more than 100 points, or by the total number of layers, particularly when three or more layers are considered (Orozco & Luque 1990).

The Q_H value for the phenols studied by all the methods was the atomic charge on the hydroxyl hydrogen. The Q_H values of amides are often incorrect because semi-empirical methods under-estimate the barrier to rotation of the NH−CO bond; a molecular mechanics correction can be added which increases the barrier (e.g. to 14 kcal mol^{-1} in N-methylacetamide; MOPAC Manual). Atomic charges calculated with or with-out these corrections were sometimes different (Table 1). The most positively charged hydrogen in the amide structures, irrespective of the method used, was always found to be the hydrogen con-nected to the nitrogen atom. (Amides with two substituents on the nitrogen were not used in

Table 1. Atomic charge and the energy of the lowest unoccupied molecular orbital (E_{LUMO}) calculated for some hydrogen-bond donors by use of the MNDO, AM1 and PM3 methods.

Compound	$\Sigma\alpha_2^H$	MNDO		AM1		PM3		PD
		Q_H	E_{LUMO}	Q_H	E_{LUMO}	Q_H	E_{LUMO}	Q_H
Ammonia	0·14	0·0756	4·3399	0·1293	4·2284	−0·0023	3·3354	–
Methylamine	0·16	0·0963	3·7070	0·1416	3·8177	0·0344	3·1062	0·31
Ethylamine	0·16	0·0940	3·4520	0·1440	3·6387	0·0414	3·0286	0·31
n-Propylamine	0·16	0·0939	3·3716	0·1441	3·5784	0·0543	2·9798	0·31
n-Butylamine	0·16	0·0944	3·2060	0·1389	3·6366	0·0595	3·0454	0·29
Dimethylamine	0·08	0·1164	3·3257	0·1522	3·4820	0·0418	2·9077	0·29
Di-n-propylamine	0·08	0·1124	2·9830	0·1511	3·1553	0·0455	2·6829	0·28
Di-n-butylamine	0·08	0·1112	2·9147	0·1525	3·1102	0·0463	2·6569	0·29
Diethylamine	0·08	0·1111	3·0364	0·1538	3·2231	0·0475	2·7364	0·29
Formamide[m]	0·62	0·1843	1·5173	0·2213	1·5696	0·0655	1·3608	0·26
Formamide[n]	0·62	0·1548	1·1539	0·2213	1·5671	0·0558	1·0770	–
Acetamide[m]	0·54	0·1844	1·3458	0·2234	1·5145	0·0660	1·2945	0·25
Acetamide[n]	0·54	0·1498	0·9794	0·2233	1·5146	0·0519	0·9558	–
Propionamide[m]	0·55	0·1847	1·3930	0·2234	1·5585	0·0663	1·3189	0·26
Propionamide[n]	0·55	0·1847	1·3923	0·2232	1·5552	0·0517	0·9625	–
N-Methylformamide[m]	0·40	0·1890	1·4263	0·2302	1·5143	0·0976	1·2696	0·25
N-Methylformamide[n]	0·40	0·1846	1·3691	0·2295	1·5047	0·0822	0·9941	–
N-Methylacetamide[m]	0·40	0·1750	1·2977	0·2223	1·5107	0·0794	1·3028	0·20
N-Methylacetamide[n]	0·40	0·1686	1·2257	0·2218	1·5055	0·0623	0·9854	–
N-Methylpropionamide[m]	0·40	0·1757	1·3458	0·2229	1·5468	0·0733	1·1673	0·21
N-Methylpropionamide[n]	0·40	0·1715	1·3119	0·2182	1·4953	0·0654	1·0278	–
Acetic acid	0·61	0·2162	0·8510	0·2429	0·9710	0·2266	0·9264	0·39
Hexanoic acid	0·60	0·2163	0·9032	0·2421	1·0158	0·2254	0·9431	0·37
Chloroacetic acid	0·74	0·2232	−0·2019	0·2489	0·1337	0·2301	0·1204	0·39
Dichloroacetic acid	0·90	0·2285	−0·6558	0·2543	−0·1576	0·2330	−0·1095	0·38
Trichloroacetic acid	0·95	0·2328	−1·0437	0·2587	−0·8124	0·2357	−0·4585	0·38
Formic acid	0·75	0·2160	0·9603	0·2417	0·9576	0·2239	0·9691	0·40
Propanoic acid	0·60	0·2165	0·9037	0·2422	1·0200	0·2258	0·9570	0·39
Butanoic acid	0·60	0·2163	0·9032	0·2422	1·0174	0·2256	0·9425	0·38
2-Methylbenzoic acid	0·60	0·2142	−0·2407	0·2447	−0·4307	0·2242	−0·2543	0·36
3-Methylbenzoic acid	0·59	0·2167	−0·5190	0·2452	−0·4266	0·2281	−0·4875	0·36
4-Methylbenzoic acid	0·60	0·2142	−0·2977	0·2458	−0·4734	0·2282	−0·5278	0·36
Water	0·82	0·1628	5·4434	0·1914	4·4180	0·1793	4·0616	–
Methanol	0·43	0·1804	3·7953	0·1954	3·7777	0·1810	3·5078	0·33
Ethanol	0·37	0·1798	3·5149	0·1965	3·5652	0·1834	3·3341	0·34
Propan-1-ol	0·37	0·1787	3·2527	0·1947	3·5987	0·1812	3·2515	0·32
Propan-2-ol	0·33	0·1780	3·3379	0·1960	3·4835	0·1840	3·2770	0·33
Butan-1-ol	0·37	0·1785	3·1939	0·1948	3·5494	0·1814	3·1732	0·33
Hexan-1-ol	0·37	0·1785	3·1495	0·1947	3·4668	0·1814	3·0812	0·32
2,2,2-Trifluoroethanol	0·57	0·2020	1·4265	0·2163	1·3804	0·1971	0·9187	0·34
Cyclopentanol	0·32	0·1792	3·0996	0·1963	3·3981	0·1839	3·2256	0·33
Cyclohexanol	0·32	0·1785	3·0662	0·1976	3·3436	0·1853	3·1481	0·33
Prop-2-en-1-ol	0·38	0·1788	0·8886	0·1982	1·0464	0·1839	0·8724	0·34
trans-But-2-en-1-ol	0·38	0·1785	0·7299	0·1967	1·0065	0·1828	0·8694	0·33
Benzylamine	0·10	0·0950	0·0564	0·1425	0·3336	0·1065	0·1428	0·31
Acetanilide[m]	0·50	0·1806	0·1568	0·2272	0·3313	0·0726	0·1715	0·25
Acetanilide[n]	0·50	0·1765	0·1602	0·2268	0·3298	0·0658	0·1592	–
Benzoic acid	0·59	0·2142	−0·2337	0·2457	−0·4721	0·2286	−0·5383	0·37
Phenol	0·60	0·1930	0·2509	0·2167	0·3998	0·1961	0·2874	0·34
2-Fluorophenol	0·61	0·2064	−0·2149	0·2296	0·0154	0·2055	−0·0619	0·35
3-Fluorophenol	0·68	0·1970	−0·2026	0·2214	0·0257	0·1996	−0·0615	0·34
4-Fluorophenol	0·63	0·1957	−0·1727	0·2198	0·0613	0·1984	−0·0478	0·34
2-Chlorophenol	0·32	0·1976	−0·1734	0·2211	0·0662	0·1995	0·0242	0·34
3-Chlorophenol	0·69	0·1968	−0·2096	0·2211	0·0377	0·1988	−0·0071	0·34
4-Chlorophenol	0·67	0·1969	−0·1799	0·2202	0·0914	0·1984	0·0518	0·34
2-Bromophenol	0·35	0·1975	−0·1204	0·2221	−0·0134	0·1955	0·0273	0·33
3-Bromophenol	0·70	0·1962	−0·1645	0·2212	−0·0486	0·2001	−0·1267	0·34
4-Bromophenol	0·67	0·1967	−0·1319	0·2216	0·0197	0·1990	−0·0241	0·35
2-Methoxyphenol	0·22	0·2069	0·1780	0·2347	0·3938	0·2074	0·3158	0·32
3-Methoxyphenol	0·59	0·1929	0·1794	0·2176	0·3628	0·1959	0·2794	0·33
4-Methoxyphenol	0·57	0·1934	0·0652	0·2175	0·2723	0·1966	0·1752	0·34
2-Nitrophenol	0·05	0·2145	−0·9596	0·2673	−1·1906	0·2598	−1·2253	0·33
3-Nitrophenol	0·79	0·2014	−0·9311	0·2255	−1·1530	0·2017	−1·1808	0·35

Table 1 continued

Table 1 continued

Compound	$\Sigma\alpha_2^H$	MNDO		AM1		PM3		PD
		Q_H	E_{LUMO}	Q_H	E_{LUMO}	Q_H	E_{LUMO}	Q_H
4-Nitrophenol	0·82	0·2035	−0·8243	0·2291	−1·0676	0·2061	−1·0824	0·35
1-Naphthol	0·61	0·1956	−0·3137	0·2200	−0·2537	0·1992	−0·3647	0·33
2-Naphthol	0·61	0·1935	−0·3943	0·2186	−0·3460	0·1971	−0·4509	0·35
Benzyl alcohol	0·33	0·1798	0·0886	0·1986	0·2837	0·1839	0·1673	0·33
Pyrrole	0·41	0·1991	1·2628	0·2415	1·3785	0·1421	1·1098	0·22

[m]With molecular mechanics correction.[n]Without molecular mechanics correction. These corrections are not relevant for the PD Q_H values.

correlations.) The most positive atomic charge in the alcohols investigated was calculated by all the methods to be on the hydroxyl hydrogen. The most positive hydrogen atom in carboxylic acids was, as expected, the acidic hydrogen.

The PM3 method underestimated the atomic charge on the nitrogen atoms and on the hydrogen atoms connected to these nitrogens, with the result that occasionally carbon atoms in the molecules were more negatively charged than nitrogen atoms. In amines the atomic charges on the hydrogens connected to the nitrogens were used as Q_H values even though sometimes in the PM3 method they were not the most positively charged in the structure.

Regression analyses between the experimental hydrogen-bonding parameter ($\Sigma\alpha_2^H$) and selected molecular orbital parameters (Q_H, E_{LUMO}) were performed among the different classes of compound listed in Table 1 (including alcohols, amines, amides, carboxylic acids, phenols and thiols) and also within the classes.

Correlations across all classes
A summary of the correlations among the different classes of compound is given in Table 2. 2-Sub-

stituted phenols and water have been deleted from all the correlations; they were outliers from the correlations except for the correlation between $\Sigma\alpha_2^H$ and PD Q_H, for which inclusion of 2-substituted phenols did not change the correlation significantly (the correlation coefficient for this equation was 0·476). In 2-substituted phenols the short distance between the phenolic hydroxyl group and appropriate substituents in the 2-position can enable intramolecular hydrogen-bonding. Substituents in this position can also give rise to steric shielding of the hydroxyl group, which limits access of the incoming lone pair. These two properties do not affect the corresponding Q_H (Mulliken) or E_{LUMO} values, with the result that the Q_H and E_{LUMO} of 2-substituted phenols are similar to those of 3- and 4-substituted phenols but their hydrogen-bond donor capacities are much lower (Table 1). Because PD charges are affected by steric factors, PD Q_H values were expected to follow the experimental $\Sigma\alpha_2^H$ values for all the substituted phenols. In correlations for the hydrogen-bond acids using the parameters calculated by the AM1 method, pyrrole was an outlier and has been omitted from regressions. Its inclusion worsened the statistics of equation 6 to $r = 0.908$ and $s = 0.105$. In the PM3 and MNDO

Table 2. Comparison of the correlations, $\sum\alpha_2^H = A(Q_H) + B(E_{LUMO}) + C$, between $\sum\alpha_2^H$ and the parameters calculated by use of different MO methods for mixed classes of compound.

Method	Parameter	Equation	A	B	C	n	r	s	F
MNDO	Q_H	1	4·86	–	−0·390	55	0·880	0·107	182·6
	E_{LUMO}	2	–	−0·113	0·618	55	0·793	0·137	89·7
	Q_H&E_{LUMO}	3	3·51	−0·051	−0·086	55	0·918	0·090	138·6
AM1	Q_H	4	5·82	–	−0·729	54	0·896	0·101	212·3
	E_{LUMO}	5	–	−0·111	0·632	54	0·796	0·138	89·5
	Q_H&E_{LUMO}	6	4·42	−0·040	−0·384	54	0·917	0·092	134·7
PM3	Q_H	7	2·45	–	0·088	55	0·759	0·147	71·9
	E_{LUMO}	8	–	−0·119	0·618	55	0·763	0·146	73·7
	Q_H&E_{LUMO}	9	1·55	−0·076	0·322	55	0·857	0·117	71·8
PD MNDO	Q_H	10	2·42	–	−0·301	54	0·507	0·192	18·0
	E_{LUMO}	11	–	0·112	0·619	54	0·791	0·137	87·0
	Q_H&E_{LUMO}	12	1·20	0·099	0·215	54	0·825	0·127	54·3

methods deletion of pyrrole did not change the equations significantly. The OML software does not accept the SMILES codes for molecules with only one non-hydrogen atom and, therefore, water and ammonia could not be included in the equations of the PD MNDO method.

It can be concluded from Table 2 that in correlations of a single parameter with $\Sigma\alpha_2^H$, MO parameters calculated by the AM1 method produced equations with the best statistics; the MNDO method was second best. It has been shown for a variety of compounds that the AM1 method is the best for modelling experimental dipole moments and ionization potentials (Stewart 1989). In multiple regression, parameters calculated by the MNDO method work best. Judging by the equations listed in Table 2, the PD Q_H values are not as good as those obtained by the Mulliken method. A plot of $\Sigma\alpha_2^H$ against PD Q_H shows that amides are outliers (Figure 1). Deleting amides from the correlations results in the equations:

$$\Sigma\alpha_2^H = 6.72Q_H - 1.80 \qquad (13)$$

$$n = 46 \quad r = 0.827 \quad s = 0.135 \quad F = 95.1$$

$$\Sigma\alpha_2^H = 4.33Q_H - 0.066E_{LUMO} - 0.904 \qquad (14)$$

$$n = 46 \quad r = 0.904 \quad s = 0.103 \quad F = 96.8$$

Inclusion of 2-substituted phenols does not worsen the correlation with PD Q_H values significantly (equation 15).

$$\Sigma\alpha_2^H = 6.90Q_H - 1.87 \qquad (15)$$
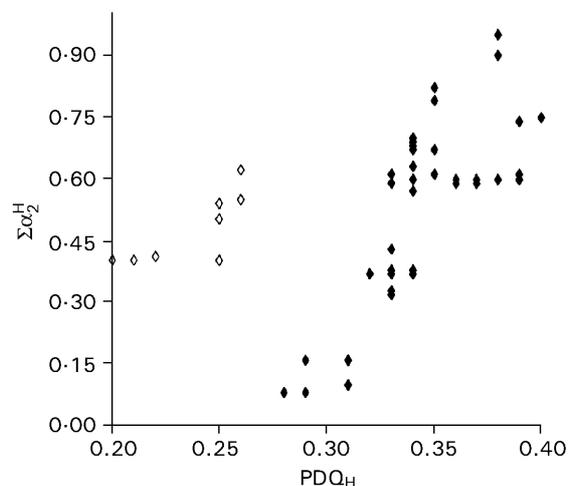
$$n = 51 \quad r = 0.809 \quad s = 0.142 \quad F = 92.5$$

Figure 1. Plot of $\Sigma\alpha_2^H$ against PD Q_H (calculated from MNDO output) for amides ($\diamond$) and for hydrogen-bonding acids ($\blacklozenge$).

This might be because the PD atomic charges are affected by steric factors and might be regarded as solvent-accessible atomic charges. Q_H and E_{LUMO} together do not correlate well when the 2-substituted phenols are included (the r value for the correlation of $\Sigma\alpha_2^H$ with Q_H and E_{LUMO} is reduced from 0.904 to 0.843).

Correlations within chemical classes

Phenols. In correlations within families of compounds, Q_H and E_{LUMO} correlate very well with each other and cannot be used in a multiple regression. The results for phenols are listed in Table 3. In this table correlations of $\Sigma\alpha_2^H$ with all the parameters are for 3- and 4-substituted phenols except for the correlation with the PD Q_H values which includes the 2-substituted phenols (omitting the 2-substituted phenols did not improve the correlation in this case). It can be concluded from Table 3 that the MNDO-calculated Q_H and E_{LUMO} are best for the prediction of the $\Sigma\alpha_2^H$ values; those calculated by use of AM1 are almost as good, and those by use of PM3 are worst.

Amides. Table 4 shows the relationships between the hydrogen-bond experimental acidity of amides with MO parameters calculated by use of different methods. As for the phenols, MNDO-calculated parameters seem to work best; this is the only method which gives Q_H values (Mulliken or PD) which correlate well with experimental hydrogen-bond parameters. In the AM1 and PM3 methods charge parameters either have an order opposite to that suggested by the inductive effect of alkyl groups in solution or do not follow any particular order. E_{LUMO} (when calculated by use of the AM1 or PM3 methods) has a positive slope in the relationships with $\Sigma\alpha_2^H$ of amides.

Alcohols. Q_H values calculated for alcohols by use of the AM1 and PM3 methods suggest that alkyl groups have an electron-withdrawing inductive effect. In these methods, if 2,2,2-trifluoroethanol is included the coefficients of Q_H have the correct signs; if this alcohol is excluded, because it is the only alcohol with a strong electron-withdrawing substituent and, therefore, its hydrogen-bonding strength is much different from those of the other alcohols, the wrong sign is obtained for the coefficients on the Q_H parameter. There is poor positive correlation between MNDO-calculated Mulliken Q_H values and $\Sigma\alpha_2^H$; the correlation improves when values for 2,2,2-trifluoroethanol are used in the equation. For PD Q_H values there

Table 3. Comparison of the correlations, $\sum \alpha_2^H = A\,(Q_H \text{ or } E_{LUMO}) + B$, between $\sum \alpha_2^H$ values for phenols and MO parameters calculated by use of different semi-empirical methods.

Method	Parameter	Equation	A	B	n	r	s
MNDO	Q_H	16	23·1	−3·87	13	0·964	0·021
	E_{LUMO}	17	−0·183	0·620	13	0·829	0·044
AM1	Q_H	18	21·4	−4·05	13	0·959	0·022
	E_{LUMO}	19	−0·127	0·648	13	0·819	0·045
PM3	Q_H	20	25·6	−4·44	13	0·920	0·031
	E_{LUMO}	21	−0·127	0·639	13	0·791	0·048
PD MNDO	Q_H	22	15·6	4·75	18	0·679	0·152

Table 4. Comparison of the correlations between $\sum \alpha_2^H$ for primary and secondary amides and MO parameters calculated by different semi-empirical methods.

Parameter		Q_H			E_{LUMO}	
	n	r	s	n	r	s
MNDO(MM[a])	6	0·942	0·033		—[b]	
MNDO	Reverse order				—[b]	
AM1(MM)		—[b]		6	0·631	0·084
AM1		—[b]		6	0·895	0·048
PM3 (MM)	Reverse order			7	0·953	0·079
PM3		—[b]			—[b]	
PD MNDO	6	0·934	0·035		—[b]	

[a]Molecular mechanics. [b]No correlation observed.

is no correlation in this class. E_{LUMO} for these structures does not correlate well with $\Sigma\alpha_2^H$ in either method; with the MNDO and PM3 methods, unexpectedly, E_{LUMO} has a positive slope in poor correlations with $\Sigma\alpha_2^H$; when the AM1 method is used there is no correlation between them. Because of the narrow range of both experimental and theoretical parameters, the inadequacy of the theoretical parameters in this particular class of compound has not affected the good correlations among different classes.

Carboxylic acids. A summary of correlation results for carboxylic acids is shown in Table 5. Mulliken Q_H values for the carboxylic acids calculated by use of these three methods correlate very well with hydrogen-bond acidity. The best method here is the MNDO method, with the AM1 method being the second best. Deletion of resonance structures improves the correlations for the PM3 and AM1 methods, but the MNDO method is still the best. PD Q_H was not successful unless the value of E_{LUMO} calculated by use of the MNDO method was also used in the correlation analyses:

$$\Sigma\alpha_2^H = 7{\cdot}55Q_H - 0{\cdot}156E_{LUMO} - 2{\cdot}16 \qquad (23)$$

$$n = 12 \quad r = 0{\cdot}865 \quad s = 0{\cdot}0716 \quad F = 13{\cdot}4$$

Table 5. Comparison of correlations between $\Sigma\alpha_2^H$ for carboxylic acids and MO parameters calculated by different semi-empirical methods.

Parameter	n	Q_H		n	E_{LUMO}	
		r	s		r	s
MNDO	12	0·934	0·048	8	0·908	0·065
AM1	12	0·858	0·069	8	0·922	0·060
PM3	12	0·744	0·086	8	0·906	0·065
PD MNDO	12	0·370	0·126			

In equation 23, t-ratios for Q_H and E_{LUMO} were 4·17 ($P < 0{\cdot}001$) and 4·67 ($P < 0{\cdot}001$) respectively. There was no correlation with carboxylic acid $\Sigma\alpha_2^H$ values of LUMO energies calculated by any of the three methods unless the aromatic structures were deleted (as in Table 5). E_{LUMO} calculated by use of the AM1 method is better than those calculated by the MNDO and PM3 methods.

Amines. There was little or no correlation with $\Sigma\alpha_2^H$ values of Q_H and E_{LUMO} values calculated by all the semi-empirical methods. For all the methods the order of the Q_H values was the opposite of that expected according to the inductive effect of alkyl groups in solution, and the gas-phase acidities of the amines were not in accord with the inductive effect of the alkyl groups (Brauman & Blair 1971).

General. Relationships between theoretical (gas phase) parameters and experimental hydrogen-bond acidities measured in solution for different families were analysed. For some compounds (amides, alcohols and amines), for which the substituents were alkyl chains only, the order of hydrogen-bonding capacity predicted by molecular orbital calculations was the opposite of those measured in solution (or the ordering was irregular); this is not in accord with the inductive electron-donor nature of alkyl groups. This has also been observed for parameters calculated by the CNDO method—Q_H values for *N*-methylamides were higher than those for the corresponding non-substituted compounds (Ghafourian 1996). It has been pointed out (Hehre & Pople 1970) that according to ab-initio (STO-3G) calculations of atomic charges a methyl group is electron-attracting (relative to hydrogen) in amines, alcohols and ethers. Despite this the order of the total energies of the neutral, protonated and deprotonated systems reproduced the experimental ordering of proton affinity. This discrepancy might be because of the effect of solvation. In a comparison between the aqueous and gas-phase acidity of a-ketoaldoximes (Bouchoux et al 1991) it was found that the global substituent effect on acidity is less in solution than in the gas phase. This can be explained partly by the non-specific weakening of the electric field effect of the substituent by the solvent permittivity (dielectric constant).

Application to QSAR

Because MO parameters such as Q_H and E_{LUMO} can model $\Sigma\alpha_2^H$ well, they should be able to serve as hydrogen-bonding parameters in quantitative structure–activity relationships. We have therefore examined how well other hydrogen-bonding parameters can be replaced in QSAR by the MO parameters considered in this work. We have selected two examples from toxicology, because environmental effects and health hazards are both legitimate concerns of the pharmaceutical industry.

Kamlet et al (1986) published QSAR studies relating the toxicities of a wide range of chemicals to the bacterium *Photobacterium phosphoreum* (now known as *Vibrio fischeri*) to their solvatochromic parameters:

$$\log EC50 = 7\cdot61 - 4\cdot11V/100 - 1\cdot54\pi^* \\ - 1\cdot51\alpha_m + 3\cdot94\beta \qquad (24)$$

$$n = 38 \quad r = 0\cdot985 \quad s = 0\cdot28$$

where EC50 is the dose resulting in the death of 50% of the bacteria, V is the molar volume, π^* is a polarity term, α_m is a term similar to $\Sigma\alpha_2^H$, and β_H is the hydrogen-bond-acceptor capacity. The correlation coefficient without the α_m and β terms included was $0\cdot876$. We replaced α_m with Q_H, and also used the charge on the most negative atom (Q_{MN}) to replace β, with the result:

$$\log EC50 = 7\cdot17 - 4\cdot25V/100 - 1\cdot16\pi^* \\ - 8\cdot87Q_H - 4\cdot85Q_{MN} + 0\cdot234E_{LUMO} \qquad (25)$$

$$n = 38 \quad r = 0\cdot979 \quad s = 0\cdot36$$

Abraham et al (1990b) reported a QSAR for irritation of the upper respiratory tract in mice by a range of non-reactive chemicals:

$$\log 1/C = -0\cdot69 + 0\cdot77\delta_2 + 2\cdot81\pi^*_2 \\ + 2\cdot82V_X + 4\cdot93\alpha_2^H \qquad (26)$$

$$n = 39 \quad r = 0\cdot985 \quad s = 0\cdot14$$

where δ_2 is a polarizability term and V_X is McGowan's characteristic volume. The correlation coefficient without the α_2^H term included was $0\cdot850$. We replaced α_2^H with Q_H to yield equation 27:

$$\log 1/C = -0\cdot70 + 0\cdot91\delta_2 + 2\cdot47\pi^*_2 \\ + 2\cdot74V_X + 13\cdot1Q_H \qquad (27)$$

$$n = 39 \quad r = 0\cdot984 \quad s = 0\cdot14$$

It is clear from these examples that Q_H and E_{LUMO}, and even Q_H alone, can model hydrogen-bond-donor capacity well in QSAR studies, and thus provide QSAR practitioners with readily calculable quantitative hydrogen-bond-donor parameters.

Conclusions

It can be concluded from these correlations that the parameters calculated by semi-empirical methods in MOPAC are useful in the prediction of experimental hydrogen-bonding capacities. The AM1 and MNDO methods give better correlations than does the PM3 method both within families and among families. One important disadvantage of the PM3 method is underestimation of negative charge on nitrogen atoms and of positive charge on hydrogen atoms connected to them (which is the reason for the poor statistics of general correlations obtained for this method).

Despite the known advantages of PD charges over Mulliken charges, and although these charges have the advantage of being affected by steric

factors, they do not correlate with experimental hydrogen-bond values as well as do Mulliken charges both among and within families. To find any relationship with E_{LUMO} within families of compounds it is always necessary to separate the conjugated structures. In correlations within families the energies of frontier orbitals and charge parameters cannot both be used because they are highly correlated. Among the different families it is essential to use both parameters to improve the prediction of hydrogen-bonding acidity and to reduce family-dependent behaviour.

The effect of alkyl substitution on charge parameters in different families is the subject of controversy. In amines, amides and alcohols these groups seem to have an electron-withdrawing effect in the AM1 and PM3 methods. The MNDO calculation of Q_H values for alcohols and amides predicts an electron-donating inductive effect for alkyl groups.

On the basis of the work reported here, it is recommended that calculation of atomic charges and E_{LUMO} for modelling hydrogen-bonding be performed by the AM1 or MNDO methods. The ability of Q_H and E_{LUMO} to model hydrogen-bond-donor capacity in QSAR has been demonstrated in two examples from toxicology.

References

Abraham, M. H. (1993) Scales of solute hydrogen-bonding. Chem. Soc. Rev. 73–83

Abraham, M. H., Grellier, P. L., Prior, D. V., Duce, P. P., Morris, J. J., Taylor, P. J. (1989) Hydrogen-bonding. Part 7. A scale of solute hydrogen-bond acidity based on K values for complexation in tetrachloromethane. J. Chem. Soc. Perkin Trans. 2: 699–711

Abraham, M. H., Grellier, P. L., Prior, D. V., Morris, J. J., Taylor, P. J. (1990a) Hydrogen-bonding. Part 10. A scale of solute hydrogen-bond basicity using log K values for complexation in tetrachloromethane. J. Chem. Soc. Perkin Trans. 2: 521–529

Abraham, M. H., Whiting, G. S., Alarie, Y., Morris, J. J., Taylor, P. J., Doherty, R. M., Taft, R. W., Nielsen, G. D. (1990b) Hydrogen-bonding. 12. A new QSAR for upper respiratory tract irritation by airborne chemicals in mice. Quant. Struct.-Act. Relat. 9: 6–10

Bouchoux, G., Jaudon, P., Decouzon, M., Gal, J.-F., Maria, P.-C. (1991) Gas-phase acidity of some α-ketoaldoximes—experiment and theory. J. Phys. Org. Chem. 4: 285–292

Brauman, J. I., Blair, L. K. (1971) Gas-phase acidities of amines. J. Am. Chem. Soc. 93: 3911–3914

Dearden, J. C., Ghafourian, T. (1993) A hydrogen-bond parameter from molecular orbital theory: comparison with α values. J. Pharm. Pharmacol. 45: 1143

Dearden, J. C., Ghafourian, T. (1995) Investigation of calculated hydrogen-bonding parameters for QSAR. In: Sanz, F., Giraldo, J., Manaut, F. (eds) QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications. Prous Science Publishers, Barcelona, pp 117–119

Fujita, T., Nishioka, T., Nakajima, M. (1977) Hydrogen-bonding parameter and its significance in quantitative structure-activity studies. J. Med. Chem. 20: 1071–1081

Ghafourian, T. (1996) Investigation of Calculated Hydrogen-bonding Parameters, PhD thesis, Liverpool John Moores University, 1996

Hehre, W. J., Pople, J. A. (1970) Methyl inductive effect on acid-base strengths. Tetrahedron Lett. 2959–2962

Kamlet, M. J., Doherty, R. M., Veith, G. D., Taft, R. W., Abraham, M. H. (1986) Solubility properties in polymers and biological media. 7. An analysis of toxicant properties that influence inhibition of bioluminescence in *Photobacterium phosphoreum* (the microtox test). Environ. Sci. Technol. 20: 690–695

Maria, P.-C., Gal, J.-F., de Franceschi, J., Fargin, E. (1987) Chemometrics of the solvent basicity—multivariate analysis of the basicity scales relevant to nonprotogenic solvents. J. Am. Chem. Soc. 109: 483–492

Orozco, M., Luque, F. J. (1990) On the use of AM1 and MNDO wave-functions to compute accurate electrostatic charges. J. Comp. Chem. 11: 909–923

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

Singh, U. C., Kollman, P. A. (1984) An approach to computing electrostatic charges for molecules. J. Comp. Chem. 5: 129–145

Stewart, J. J. P. (1989) Optimization of parameters for semi-empirical methods II. Applications. J. Comp. Chem. 10: 221–264

Stewart, J. J. P. (1990) MOPAC—a semi-empirical molecular orbital program. J. Computer-Aided Mol. Design 4: 1–105

Wilson, L. Y., Famini, G. R. (1991) Using theoretical descriptors in quantitative structure-activity relationships—some toxicological indexes. J. Med. Chem. 34: 1668–1674