Prediction of Hydrogen Bond Basicity from Computed Molecular Electrostatic Properties: Implications for Comparative Molecular Field Analysis

Peter W. Kenny
Zeneca Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire, UK SK10 4TG

Computed molecular electrostatic properties have been evaluated as predictors of hydrogen bond basicity for a set of heterocycles with nitrogen as the hydrogen bond acceptor. The properties were (*a*) the electrostatic potential local minimum, V_{min}, in the region of the nitrogen lone pair, (*b*) the electrostatic potential, $V_\beta(r)$, and (*c*) the magnitude of the electric field strength, $|\boldsymbol{F}_\beta(r)|$, at points along the lone pair axis, defined by the distance r from nitrogen. V_{min} was found to be an excellent predictor of hydrogen bond basicity. The ability of the other parameters to fit the data was maximised at r = 1.4 Å for $V_\beta(r)$ and at r = 2.5 Å for $|\boldsymbol{F}_\beta(r)|$ which has implications for Comparative Molecular Field Analysis.

Hydrogen bonding is a key element of biomolecular recognition, being implicated in DNA base pairing, protein folding and enzymatic catalysis.[1,2] Specifically, hydrogen bonding is an important determinant of the strength of binding of drug molecules to their targets, and the ability to make quantitative predictions is of value in medicinal chemistry.[3]

Hydrogen bonding is usually considered essentially electrostatic in nature, which is one rationale for using atomic charges as descriptors for quantitative structure activity relationships (QSAR). Calculation of atomic charges partitions the total electronic charge of a molecule and there are various ways in which this can be done ranging from the computationally simple Mulliken population analysis[4] to a number of methods for fitting charges to the electrostatic field around a molecule.[5,6] A legitimate concern with using atomic charges as descriptors is that differences in atomic charges may be a function of the partitioning scheme as well as of the electrical properties of the atoms concerned. Furthermore, there are limits to how accurately atomic charges can reproduce molecular electrostatics, although the inclusion of multipoles improves the situation.[7,8]

An alternative strategy for deriving electrostatic descriptors is to focus on the electrostatic field around the molecule, rather than atomic charges, in the process eliminating the need to match atoms in the overlays. Comparative Molecular Field Analysis (CoMFA) is a QSAR technique which escapes from the constraints of an atomic paradigm by exploiting the electrostatic and steric fields around molecules.[9] The molecular electrostatic field is sampled by computing electrostatic potential on a lattice around the molecule. Atomic charges from a number of sources are used to calculate the electrostatic potentials, although there is no reason why the electrostatic potentials could not be obtained directly from a wavefunction.

There is precedent for using calculated electrostatic potentials in the quantitative description of hydrogen bonding.[10-12] Acceptors were characterised by electrostatic potential local minima (V_{min}) and correlations with the relevant solute hydrogen bonding parameters were presented. Note that is is necessary to compute electrostatic potential directly from the wavefunction for these minima to be observed.

This study first extends the work of these authors to a set of 23 aromatic heterocycles and shows that useful predictions can also be made for heterocycles with non-equivalent nitrogen acceptors. However, the principal conclusion of this work is that the gradient of the electrostatic potential, the electric field, is a useful descriptor of hydrogen bond basicity.

Computational Details

All quantum mechanical calculations were performed with the GAUSSIAN 88 electronic structure program on a Convex C220 computer.[13] Molecular structures were energy-minimised using the 3-21G(*) basis set[14] and electrostatic properties were calculated directly from the wavefunction using the 6-31G* basis set.[15] The lone pair axis of each nitrogen acceptor was defined geometrically by the exocyclic segment of the line of intersection between the least squares plane of the ring and the plane bisecting the angle at nitrogen. Electrostatic potential minimisations were initiated 1.3 Å from nitrogen on the lone pair axis while computation of electric field $\boldsymbol{F}_\beta(r)$ and electrostatic potential $V_\beta(r)$ were carried out at points along this axis at specified distances, r, from nitrogen. Tables of values of $\boldsymbol{F}_\beta(r)$ and $V_\beta(r)$ as r varies for 23 compounds are available as supplementary data.† Experimental hydrogen bond basicities[16] (log K_β, donor: 4-nitrophenol, solvent: 1,1,1-trichloroethane) were corrected statistically when two equivalent nitrogen atoms were present. All statistical analyses were carried out with the SAS software package.[17]

Results and Discussion

The rationalisation of the hydrogen bond basicities of heterocycles is an interesting problem with relevance to drug design. It is well known that pK_a is not a particularly good predictor of hydrogen bond basicity; the observation that pyridine and pyridazine are equally strong hydrogen bond acceptors, despite a three unit difference in pK_a, is a case in point.[16] Table 1 lists V_{min} and log K_β for heterocycles which have either a single nitrogen acceptor or equivalent nitrogen acceptors. The predicted log K_β values are given by eqn. (1). The

$$\log K_\beta = -5.125 - 2.533 \times 10^{-2} (V_{min}/\text{kJ mol}^{-1}) \quad (1)$$

$$(R^2 = 0.963, s = 0.160, F = 553, n = 23)$$

agreement between predicted hydrogen bond basicities and the corresponding experimental values is excellent over three orders of magnitude and the largest residual is only −0.33. Heterocycles such as **13**, **20** and **21** with lone pairs on adjacent

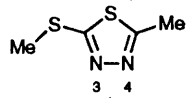
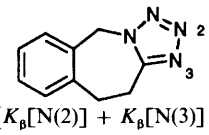
† For details of the supplementary publication scheme, see 'Instructions for Authors', *J. Chem. Soc., Perkin Trans. 2*, 1994, issue 1 (Supp. Pub. no. 56983, 5 pp).

Table 1 Minimised electrostatic potentials and hydrogen bond basicities

			log K_β		
		V_{min}/kJ mol^{-1}	Expt[a]	Pred[b]	Resid[c]
1	Pyridine	−301.7	2.52	2.51	0.01
2	2-Fluoropyridine	−264.8	1.41	1.58	−0.17
3	3-Fluoropyridine	−270.5	1.82	1.72	0.10
4	2-Chloropyridine	−263.6	1.48	1.55	−0.07
5	3-Chloropyridine	−266.2	1.77	1.62	0.15
6	3-Methylpyridine	−305.8	2.65	2.62	0.03
7	3,4-Dimethylpyridine	−314.5	3.06	2.84	0.22
8	4-Methylpyridine	−310.9	2.78	2.75	0.03
9	4-Methoxypyridine	−315.0	2.87	2.85	0.02
10	4-*N*,*N*-Dimethylaminopyridine	−343.6	3.54	3.58	−0.04
11	Pyrazine	−253.6	1.16	1.30	−0.14
12	Pyrimidine	−269.3	1.37	1.69	−0.33
13	Pyridazine	−292.3	2.23	2.28	−0.05
14	Oxazole	−277.8	1.67	1.91	−0.24
15	Isoxazole	−245.7	1.06	1.10	−0.04
16	Thiazole	−273.3	1.90	1.80	0.10
17	1-Methylpyrazole	−292.0	2.22	2.27	−0.05
18	1-Methylimidazole	−348.2	3.68	3.69	−0.01
19	2,4,5-Trimethyloxazole	−304.5	2.65	2.59	0.06
20	1-Butyl-1,3,4-triazole	−330.7	3.07	3.25	−0.18
21	2,5-dimethylthia-3,4-diazole	−278.1	2.21	1.92	0.29
22	Benzothiazole	−259.4	1.76	1.45	0.31
23	Benzothia-2,5-diazole	−221.8	0.49	0.49	0.00

[a] Expt = measured hydrogen bond basicity from ref. 16. [b] Pred = hydrogen bond basicity predicted from eqn. (1). [c] Resid = Expt − Pred.

Table 2 Predicted hydrogen bond basicities for heterocycles with two or more non-equivalent nitrogen acceptors

				log K_β		
			V_{min}/kJ mol^{-1}	Pred[a]	Expt[b]	Resid[c]
24	1-Benzyl-1,2,4-triazole	N(2)	−242.1	1.01		
		N(4)	−306.2	2.63		
	log {K_β[N(2)] + K_β[N(4)]}			2.64	2.38	−0.26
25		N(3)	−248.1	1.16		
		N(4)	−262.9	1.53		
	log {K_β[N(3)] + K_β[N(4)]}			1.68	1.98	0.30
26	1-Phenethyl-1,2,3-triazole	N(2)	−239.4	0.94		
		N(3)	−301.0	2.50		
	log {K_β[N(2)] + K_β[N(3)]}			2.51	2.36	−0.17
27	1-Methylbenzotriazole	N(2)	−217.8	0.40		
		N(3)	−277.1	1.89		
	log {K_β[N(2)] + K_β[N(4)]}			1.90	2.17	0.27
28		N(1)	−204.8	0.06		
		N(2)	−274.1	1.82		
		N(3)	−296.8	2.39		
	log {K_β[N(2)] + K_β[N(3)] + K_β[N(4)]}			2.50	1.99	−0.51

[a] Pred = log K_β predicted from eqn. (1). [b] Expt = experimental log K_β from ref. 16. [c] Resid = Expt − Pred.

atoms no longer appear to be unusually strong hydrogen bond acceptors and it is not necessary to use separate equations for five- and six-membered rings as is the case when fitting log K_β to pK_a.

These calculations can also be used to predict the hydrogen bond basicities for heterocycles with two or more inequivalent nitrogen acceptors. In this case, V_{min} is computed for each acceptor nitrogen and the K_β values predicted by eqn. (1) are summed to give a prediction for the overall hydrogen bond basicity. The results of these calculations for heterocycles **24**–**28** are given in Table 2. Once again, agreement with experiment is good, with predicted and measured log K_β values falling within 0.30 of each other for all compounds except the tetrazole, **28**, for which the prediction is too high by 0.51.

While relevant to medicinal chemistry, the results presented in Tables 1 and 2 have little bearing on CoMFA because lattice points do not in general correspond to electrostatic potential minima. While the strong correlations between log K_β and V_{min} do suggest molecular alignment by means of pseudo-atoms at electrostatic potential minima, the key issue in making the connection with CoMFA is how effectively electrostatic potential predicts hydrogen bond basicity when it is computed at points away from local minima. A lattice spacing 2.0 Å is typical, implying that a point within the lattice could be up to 1.73 Å from the nearest vertex. There is some debate[18,19] as to whether a finer lattice leads to better predictions although an insensitivity to spacing may be symptomatic of non-optimum alignment of the molecules in question.

Table 3 The fit of eqn. (2) to the hydrogen bonding data of ref. 16 as a function of r; standard errors for regression coefficients have been indicated

r/Å	k_2	$v_2(r)/10^{-2}$ kJ^{-1} mol	R^2	s	F
1.00	−2.047 ± 0.249	−2.031 ± 0.118	0.933	0.216	294
1.10	−3.611 ± 0.314	−2.182 ± 0.118	0.942	0.202	341
1.20	−4.523 ± 0.341	−2.350 ± 0.119	0.949	0.190	388
1.30	−5.014 ± 0.351	−2.531 ± 0.123	0.952	0.182	421
1.40	−5.222 ± 0.359	−2.720 ± 0.131	0.953	0.182	426
1.50	−5.231 ± 0.373	−2.914 ± 0.146	0.950	0.188	396
1.60	−5.100 ± 0.394	−3.108 ± 0.168	0.942	0.202	341
1.70	−4.871 ± 0.422	−3.296 ± 0.197	0.930	0.221	280
1.80	−4.577 ± 0.452	−3.476 ± 0.232	0.914	0.245	224
1.90	−4.248 ± 0.480	−3.645 ± 0.272	0.895	0.271	180
2.00	−3.904 ± 0.505	−3.803 ± 0.315	0.874	0.297	146
2.10	−3.562 ± 0.525	−3.951 ± 0.360	0.851	0.323	120
2.20	−3.231 ± 0.541	−4.088 ± 0.407	0.828	0.348	101
2.30	−2.919 ± 0.551	−4.218 ± 0.455	0.804	0.371	86
2.40	−2.630 ± 0.559	−4.342 ± 0.502	0.781	0.392	75
2.50	−2.363 ± 0.562	−4.461 ± 0.550	0.768	0.412	66
2.60	−2.119 ± 0.564	−4.577 ± 0.597	0.737	0.430	59
2.70	−1.897 ± 0.563	−4.691 ± 0.644	0.716	0.446	53
2.80	−1.695 ± 0.561	−4.805 ± 0.691	0.697	0.461	48
2.90	−1.511 ± 0.557	−4.919 ± 0.738	0.679	0.475	44
3.00	−1.343 ± 0.553	−5.033 ± 0.784	0.662	0.487	42
3.50	−0.702 ± 0.562	−5.635 ± 1.017	0.594	0.534	31

Table 4 Fit of eqn. (3) to the hydrogen bonding data of ref. 16 as a function of r; standard errors for regression coefficients have been indicated

r/Å	k_3	$f_3(r)/10^{-10}$ V^{-1} m	R^2	s	F
1.00	17.959 ± 3.183	−1.808 ± 0.363	0.541	0.568	25
1.10	8.590 ± 1.294	−1.758 ± 0.352	0.543	0.566	25
1.20	3.371 ± 0.322	−1.453 ± 0.350	0.451	0.621	17
1.30	−0.453 ± 0.478	2.952 ± 0.528	0.598	0.531	31
1.40	−3.456 ± 0.596	3.397 ± 0.359	0.810	0.365	90
1.50	−4.784 ± 0.761	3.437 ± 0.376	0.800	0.375	84
1.60	−5.651 ± 0.851	3.666 ± 0.399	0.801	0.374	85
1.70	−6.246 ± 0.881	4.024 ± 0.421	0.813	0.362	91
1.80	−6.655 ± 0.867	4.493 ± 0.441	0.832	0.344	104
1.90	−6.929 ± 0.820	5.065 ± 0.456	0.854	0.320	123
2.00	−7.092 ± 0.751	5.734 ± 0.464	0.879	0.292	152
2.10	−7.158 ± 0.668	6.490 ± 0.464	0.903	0.261	196
2.20	−7.137 ± 0.581	7.317 ± 0.456	0.924	0.230	257
2.30	−7.028 ± 0.503	8.190 ± 0.448	0.941	0.204	334
2.40	−6.847 ± 0.446	9.090 ± 0.449	0.951	0.185	410
2.50	−6.600 ± 0.417	9.990 ± 0.474	0.955	0.178	444
2.60	−6.301 ± 0.418	10.87 ± 0.54	0.952	0.184	412
2.70	−5.966 ± 0.441	11.72 ± 0.63	0.942	0.202	342
2.80	−5.606 ± 0.476	12.54 ± 0.77	0.927	0.226	267
2.90	−5.239 ± 0.515	13.31 ± 0.92	0.908	0.254	208
3.00	−4.866 ± 0.552	14.04 ± 1.10	0.886	0.283	163
3.50	−3.193 ± 0.662	17.20 ± 2.11	0.759	0.411	66

Table 5 Fit of eqn. (4) to hydrogen bonding data of ref. 16 as a function of r; standard errors for regression coefficients have been indicated

r/Å	k_4	$v_4(r)/10^{-2}$ kJ^{-1} mol	$f_4(r)/10^{-11}$ V^{-1} m	R^2	s	F
1.00	−9.062 ± 2.221	−2.525 ± 0.185	6.852 ± 2.159	0.956	0.181	215
1.10	−6.680 ± 1.164	−2.602 ± 0.186	5.355 ± 1.974	0.958	0.177	226
1.20	−5.566 ± 0.581	−2.622 ± 0.168	3.230 ± 1.509	0.958	0.176	229
1.30	−4.722 ± 0.332	−2.244 ± 0.157	5.908 ± 2.309	0.964	0.162	270
1.40	−5.146 ± 0.349	−2.324 ± 0.275	6.043 ± 3.723	0.959	0.175	231
1.50	−5.443 ± 0.366	−2.432 ± 0.281	7.010 ± 3.609	0.958	0.176	227
1.60	−5.667 ± 0.401	−2.442 ± 0.283	9.959 ± 3.616	0.958	0.176	228
1.70	−5.859 ± 0.429	−2.406 ± 0.288	13.83 ± 3.76	0.958	0.175	230
1.80	−6.020 ± 0.447	−2.332 ± 0.297	18.66 ± 4.02	0.959	0.174	233
1.90	−6.149 ± 0.457	−2.217 ± 0.308	24.60 ± 4.38	0.959	0.173	236
2.00	−6.248 ± 0.462	−2.055 ± 0.323	31.81 ± 4.86	0.960	0.172	239
2.10	−6.321 ± 0.464	−1.841 ± 0.343	40.50 ± 5.46	0.960	0.171	242
2.20	−6.376 ± 0.467	−1.563 ± 0.368	50.96 ± 6.23	0.960	0.171	242
2.30	−6.408 ± 0.472	−1.222 ± 0.401	63.25 ± 7.19	0.960	0.172	238
2.40	−6.427 ± 0.481	−0.810 ± 0.444	77.64 ± 8.41	0.958	0.175	230
2.50	−6.432 ± 0.496	−0.323 ± 0.499	94.26 ± 9.97	0.956	0.181	216
2.60	−6.425 ± 0.517	0.243 ± 0.572	113.2 ± 12.0	0.952	0.188	198
2.70	−6.409 ± 0.545	0.888 ± 0.664	134.7 ± 14.5	0.947	0.198	178
2.80	−6.381 ± 0.580	1.612 ± 0.779	158.7 ± 17.6	0.940	0.210	156
2.90	−6.346 ± 0.621	2.417 ± 0.921	185.5 ± 21.6	0.932	0.224	136
3.00	−6.291 ± 0.668	3.288 ± 1.096	214.5 ± 26.4	0.921	0.241	117
3.50	−5.716 ± 0.961	8.263 ± 2.577	387.7 ± 69.5	0.841	0.342	53

These issues may be addressed by systematically varying the position at which electrostatic potential is computed and observing the variation of the fit to the experimental data. The nitrogen lone pair axis provides the obvious frame of reference and the descriptor $V_\beta(r)$ was defined as the electrostatic potential on this axis at a distance r from nitrogen. The ability of eqn. (2) to fit the data is presented as a function of r in

$$\log K_\beta = k_2(r) + v_2(r)V_\beta(r) \qquad (2)$$

Table 3. The electric field strength is easily obtained from the wave function and calculations analogous to those summarised in Table 3 were carried out. The electric field, $\boldsymbol{F}_\beta(r)$, at a distance r from nitrogen along the lone pair axis, is a vector with three components which depend on the coordinate system used to describe the molecular geometry and so its magnitude, $|\boldsymbol{F}_\beta(r)|$ is a more appropriate electrostatic descriptor for the purposes of this study. In QSAR work, molecules will generally be overlaid and the individual components of the field will then be valid descriptors. Table 4 summarises the dependence on r of the ability of eqn. (3) to fit the hydrogen bonding data.

$$\log K_\beta = k_3(r) + f_3(r)|\boldsymbol{F}_\beta(r)| \qquad (3)$$

The results presented in Tables 3 and 4 show that, given the appropriate choice of r, either $V_\beta(r)$ or $|\boldsymbol{F}_\beta(r)|$ can fit the experimental hydrogen bonding data almost as effectively as V_{min}. The values of r corresponding to V_{min} (ranging from 1.21 Å to 1.28 Å) and to optimal fit for $V_\beta(r)$ (1.4 Å) are comparable but differ significantly from the value of r corresponding to optimal fit for $|\boldsymbol{F}_\beta(r)|$ (2.5 Å). Particularly relevant is the observation that electrostatic potential fits log K_β most

effectively when it is calculated within the van der Waals radius of nitrogen.[20] CoMFA generally ignores the variation in the molecular fields if the steric energy exceeds a cutoff, typically set to 125 kJ mol^{-1} (30 kcal mol^{-1}). With the general purpose TRIPOS 5.2 force field,[21] this corresponds to a separation of 2.0 Å between sp^2 nitrogen and the commonly used sp^3 carbon probe. Had CoMFA been applied to this set of hydrogen bonding data using a 1.7 Å probe, the points at which electrostatic potential best fitted the experimental hydrogen bond basicity would have been ignored.

The observation that the values of r corresponding to optimal fit are significantly different for $V_\beta(r)$ and $|F_\beta(r)|$ suggests the use of both descriptors to fit the data. The results of fitting the data with eqn. (4) are presented in Table 5. The fit to log K_β is signifi-

$$\log K_\beta = k_4 + v_4(r)V_\beta(r) + f_4(r)|F_\beta(r)| \qquad (4)$$

cantly less sensitive to r than when either descriptor is used alone; R^2 only decreases from 0.964 to 0.952 as r increases from 1.30 Å to 2.60 Å. This result implies that using both potential and its gradient, the electric field will provide a superior description of the molecular electrostatic field. Using both electric field and electrostatic potential will have a similar effect to increasing the lattice resolution.

The importance of CoMFA in QSAR research is that it eliminates the need for atom-based descriptors by coupling molecular modelling methodology with modern multivariate analysis. This study suggests that electrostatic potential is typically not sampled closely enough to hydrogen bond acceptor atoms and perhaps the radius of the probe should be made to depend on whether the interaction is with an atom capable of hydrogen bonding. At greater distances from the acceptor, the electric field strength becomes a more effective descriptor of hydrogen bond basicity than electrostatic potential. Used together, these electrostatic properties will provide a better defined molecular electrostatic field for a given lattice spacing, which may lead to improved models for biological activity.

References

1 L. Stryer, *Biochemistry*, 3rd edn., 1991, W. H. Freeman and Co., New York.
2 A. Fersht, *Enzyme Structure and Mechanism*, 2nd edn., 1985, W. H. Freeman and Co., New York.
3 J. J. Morris, L. R. Hughes, A. T. Glen and P. J. Taylor, *J. Med. Chem.*, 1991, **34**, 447.
4 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833.
5 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129.
6 L. E. Chirlian and M. M. Francl, *J. Comput. Chem.*, 1987, **8**, 894.
7 A. J. Stone, *Chem. Phys. Lett.*, 1983, **83**, 233.
8 D. E. Williams, *J. Comput. Chem.*, 1988, **9**, 745.
9 R. D. Cramer, III, D. E. Patterson and J. D. Bunce, *J. Am. Chem. Soc.*, 1988, **110**, 5959.
10 J. S. Murray, S. Raganathan and P. Politzer, *J. Org. Chem.*, 1991, **56**, 3754.
11 J. S. Murray, T. Brinck, M. E. Grice and P. Politzer, *Theochem.*, 1992, **88**, 29.
12 J. S. Murray and P. Politzer, *J. Chem. Res. (S)*, 1992, 110.
13 M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. Martin, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol and J. A. Pople, *Gaussian Inc.*, Pittsburgh, PA, 1988.
14 W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. Defrees, J. A. Pople and J. S. Binkley, *J. Am. Chem. Soc.*, 1982, **104**, 5039.
15 P. C. Harihan and J. A. Pople, *Chem. Phys. Lett.*, 1972, **66**, 217
16 M. H. Abraham, P. P. Duce, D. V. Prior, D. G. Barrett, J. J. Morris and P. J. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1989, 1355.
17 SAS Institute Inc., SAS Circle, Box 8000, Cary, NC 27512-8000.
18 G. Klebe and U. Abraham, *J. Med. Chem.*, 1993, **36**, 70.
19 G. Greco, E. Novellino, C. Silipo and A. Vittoria, *Quant. Struct.–Act. Relat.*, 1991, **10**, 289.
20 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441.
21 M. Clark, R. D. Cramer, III and N. Van Opdenbosch, *J. Comput. Chem.*, 1989, **10**, 982.

Paper 3/05316D
Received 6*th September* 1993
Accepted 14*th October* 1993