Hydrogen Bonding, Electrostatic Potential, and Molecular Design

Peter W. Kenny*

AstraZeneca, Alderley Park, Macclesfield SK10 4TG, U.K.

Received January 17, 2009

The $V_\alpha(r)$ descriptor was introduced and shown to be an effective and useful predictor of hydrogen bond acidity. $V_\alpha(r)$ is defined as the electrostatic potential at a distance, r, from the donor hydrogen on an axis defined by the nuclei of the hydrogen atom and the atom to which it is bonded. $V_\alpha(r)$ is most predictive of hydrogen bond acidity for $r = 0.55$ Å which is less than half the van der Waals radius of hydrogen. Calculated values of $V_\alpha(r)$ and minimized electrostatic potential (V_{min}) were used to show how molecular electrostatic potential can be used to provide insight into a number of hydrogen bonding phenomena, including lactam self-association, DNA base pairing, and bioisosterism. The effects of hydrogen bond formation on the strengths of other donors in the interacting molecules were explored and quantified. Implications of these results for modeling hydrogen bond acidity, derivation of atomic charges, and development of polarizable force fields were discussed.

INTRODUCTION

Hydrogen bonding[1−3] is a key element of molecular recognition and is implicated in a diverse range of physicochemical phenomena including crystal packing,[4−6] DNA base-pairing,[7,8] protein-folding,[9−14] and enzymatic catalysis.[15,16] The cohesiveness of liquid water that drives hydrophobic association[17−20] in aqueous media is a consequence of strong, cooperative hydrogen bonds between water molecules. In the pharmaceutical context, hydrogen bonding influences solubility of drugs, both in water and lipid, and the avidity with which they bind to their targets. Hydrogen bond strength can be quantified by measuring association constants for donor−acceptor complexes in nonpolar solvents, using prototypical model compounds.[21−25] The difference between the 1-octanol/water and alkane/water partition coefficients[26−28] (ΔlogP) reflects hydrogen bonding between solute molecules and water-saturated 1-octanol and has been identified as a target for predictive modeling.[29]

The assumption of equivalence of hydrogen bonds is a recurring theme in medicinal chemistry. The best example of this is the influential 'rule of 5' that warns of potential for poor oral absorption if the numbers of hydrogen bond acceptors (oxygen or nitrogen atoms) exceeds ten or donors (oxygen or nitrogen atoms bonded to one or more hydrogen atoms) exceeds 5.[30] Polar surface area (PSA),[31] commonly used as a molecular descriptor in modeling intestinal permeability and penetration into the central nervous system (CNS), takes no account of variation in hydrogen bonding potential for individual donors and acceptors. A number of the scoring functions[32] used in virtual screening[33] use the sum of intermolecular hydrogen bonds in each docked pose as a descriptor to identify compounds most likely to bind effectively to the target protein.

A number of approaches have been described for prediction of hydrogen bond strength. Outside narrowly defined structural series, pK_a is not a particularly effective predictor of hydrogen bond acidity or basicity.[25] Hydrogen bonding is largely an electrostatic phenomenon, which is a rationale for using atomic charges as descriptors in quantitative structure activity/property relationships (QSAR/QSPR). Atomic charges cannot in general be defined uniquely, and it is actually the electrostatic field around a molecule that is more directly relevant to molecular recognition.[34] This is the basis of Comparative Molecular Field Analysis (CoMFA), a widely adopted QSAR technique.[35] Minimized electrostatic potential (V_{min}) has been shown[36,37] to be a useful predictor of hydrogen bond acceptor strength, but analogous electrostatic potential maxima are not found in the neighborhoods of hydrogen bond donors. This issue has been addressed by locating electrostatic potential maxima on molecular surfaces defined using van der Waals radii[38,39] or electron density[40] although it has not actually been demonstrated that these surfaces are especially relevant to recognition of hydrogen bond donors.

Although hydrogen bond strengths can also be quantified by using quantum mechanical methods to calculate energies of the donor, acceptor, and complex, there are still significant advantages to using molecular electrostatic potential to model these. In some molecular design situations (e.g., binding to protein of unknown structure) interaction partners of hydrogen bonding groups will not be known. Even when the interaction partner is known, using models of the complex to derive hydrogen bond strengths is complicated by basis set superposition error (BSSE).[41,42] Molecular recognition in biological systems occurs in buffered aqueous media, and the strength of a hydrogen bond in gas phase is not generally predictive of association thermodynamics under physiological conditions. Kangas and Tidor describe association in aqueous systems as 'essentially an exchange reaction in which interactions with solvent in the unbound state are exchanged for intermolecular interactions between partners in the bound complex'.[43] Continuum models can be used to show that 'there is an intermediate ligand charge distribution for which

* Corresponding author phone: +44 1625 514396; fax: +44 1625 519749; e-mail: pwk.pub.2008@gmail.com.

10.1021/ci9000234 CCC: $40.75
Published on Web 04/21/2009

the electrostatic contribution to the binding free energy is as favorable as possible for a given geometry'.[43] Casting this in a hydrogen bonding framework suggests that effective interaction between a donor and acceptor in aqueous media requires that their hydrogen bonding characteristics be balanced.[39] Another factor that needs to be considered when viewing binding as an exchange reaction is that the constraints imposed by multiple contacts mean that hydrogen bonds between protein and ligand are more likely to be of nonideal geometry than hydrogen bonds between unbound protein or ligand with water. In these situations, knowledge of the individual hydrogen bonding characteristics of the donor and acceptor may actually be more useful than the equilibrium constant measured for their association in a nonpolar solvent.

Molecular design can be defined as control of properties of compounds and materials through manipulation of molecular properties. One strategy, which may be termed 'prediction-driven design', is based on the assumption that predictive models can be built for the properties of interest. In many situations, properties of compounds simply cannot be predicted with the accuracy required for meaningful design, especially when optimization is performed against multiple end points. Molecular design frequently starts with one or more compounds with suboptimal properties showing a degree of the desired functional behavior. In this situation one can pose hypotheses that variation of particular molecular properties will result in changes in the functional behavior of compounds. This approach, commonly adopted in pharmaceutical and agrochemical research, can be described as 'hypothesis-driven design'. The functional behavior of a compound is determined entirely by the interactions of its molecules with the different environments in which they exist, and hydrogen bonding is a key determinant of the nature and strength of these interactions in many molecular design scenarios. While an improved description of hydrogen bonding might be expected to allow properties of compounds and materials to be modeled more accurately, it can also provide physicochemical insight that can be used to frame design hypotheses.

This article first introduces the $V_\alpha(r)$ descriptor that can be used to predict hydrogen bond acidity just as V_{min} can be used as a predictor of hydrogen bond basicity. A number of examples will be presented to illustrate how these descriptors can be used in molecular design.

THEORETICAL CALCULATIONS

Association constants (K_α) for formation of hydrogen bonded complexes of donors with N-methylpyrrolidone in 1,1,1-trichloroethane were taken from the literature.[25,44,45] Quantum mechanical calculations were performed using the Gaussian 03 electronic structure program,[46] and all values of electrostatic potential are quoted in atomic units of negative Hartree per electron. Molecular geometries were energy-minimized using the Hartree−Fock[47] (HF) electronic structure model and the 6-31G* [48,49] basis set. The molecular geometries of the structures in the training set were also energy-minimized using this basis set and the Becke[50] three parameter exchange functional, combined with the Lee−Yang−Parr[51] correlation functional (B3LYP), to test the sensitivity of the analysis of the training set to the theoretical model used for energy-minimization. The $V_\alpha(r)$ descriptor was defined as the molecular electrostatic potential (MEP) calculated at a distance, r, from the donor hydrogen on the extension of the axis defined by the hydrogen and the atom to which it is bonded (Figure 1). Values of $V_\alpha(r)$ were calculated ($0.20 \leq r/\text{Å} \leq 4.00$ at 0.05 Å increments) for the structures in the training set using the HF, B3LYP, and MP2[52,53] (second order Møller−Plesset perturbation theory) theoretical models with the 6-31G*,[48,49] 6-31G**,[48,49] 6-31+G**,[54,49] or 6-311G**[55,49] basis sets.

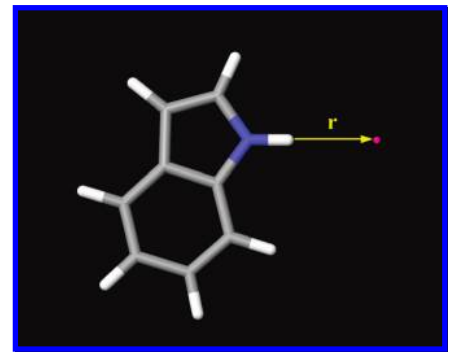


Figure 1. Location of point at which $V_\alpha(r)$ is calculated for indole.

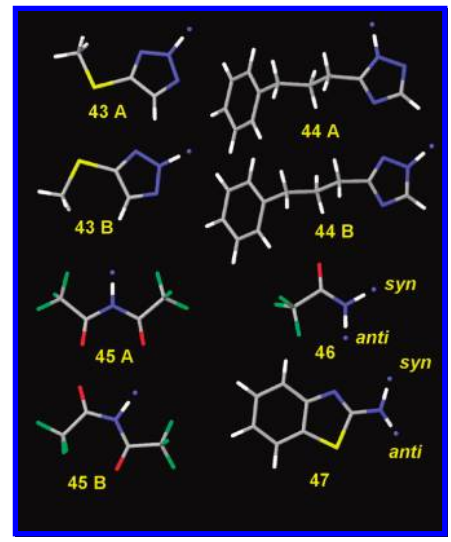


Figure 2. Validation set (see Chart 1) compounds that were treated as existing in multiple conformational (**43**, **45**) or tautomeric (**44**) states or had nonequivalent hydrogen bond donors (**46**, **47**) showing points (in blue) at which MEP was calculated as a predictor of hydrogen bond donor strength.

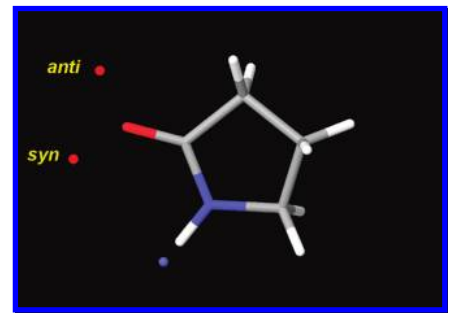
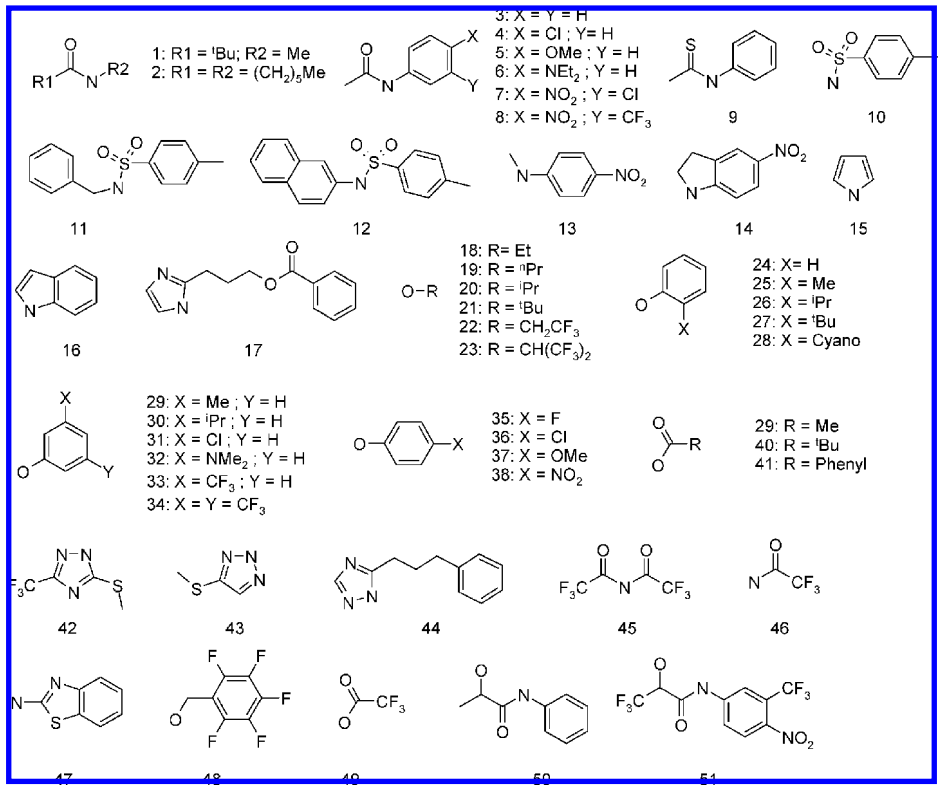


Figure 3. Electrostatic potential minima (red) and points (blue) at which $V_\alpha(r)$ was calculated to quantify effects of ring size on self-association of lactams (see Table 4). Minima are labeled *syn* or *anti* according to the geometric relationship with respect to the donor hydrogen.

Chart 1. Structures Used To Derive and Validate Relationships between Hydrogen Bond Acidity and Computed MEP



Electrostatic potential minimizations were started from points 1.2 Å from acceptor atom nuclei on geometrically defined lone pair axes using the Gaussian keyword combination Prop = (Opt,EFG).[37] The lone pair axis for each doubly connected nitrogen atom was defined by first locating a point at a fixed distance from the nitrogen atom on each of the two bond axes and then locating the centroid of these two points. The coordinates of the centroid and nitrogen atom were then used to define the lone pair axis. Lone pair axes for each carbonyl oxygen atom were defined by first defining the unit vector, **i**, by dividing the carbon oxygen double bond vector by its magnitude. The unit vector **j** was constructed from **i** and the vectors $\mathbf{b_1}$ and $\mathbf{b_2}$ for the two single bonds to the carbon atom of the carbonyl group by scaling the vector triple product of these vectors by its magnitude:

$$\mathbf{j} = \mathbf{i} \times (\mathbf{b_1} \times \mathbf{b_2})/|\mathbf{i} \times (\mathbf{b_1} \times \mathbf{b_2})| \quad (1)$$

The lone pair axes were defined by solving eq 2 in which $(180°-\alpha)$ is the angle between lone pair and carbonyl bond vectors:

$$\mathbf{a} = \mathbf{i} \cos \alpha \pm \mathbf{j} \sin \alpha \quad (2)$$

Measured values of log K_α were fit to $V_\alpha(r)$ (eq 3) using the JMP[56] program for the training set.

$$\log K_\alpha = \log K_{\alpha,0} + v(r) \times V_\alpha(r) \quad (3)$$

Quality of fit was quantified as a function of r, theoretical model and basis set by root-mean-square error (RMSE), R-square, and F-ratio. The best fit to the experimental data for the training set was found for $r = 0.55$ Å, using the B3LYP theoretical model with the 6-31+G** basis set and molecular geometries that had been energy-minimized at the RHF/6-31G* level. All subsequent calculations described in this study used this protocol, and the energy-minimized geometries are provided as Supporting Information.

RESULTS AND DISCUSSION

Relationships between Hydrogen Bond Acidity and Calculated MEP. The first objective of this study was to derive an electrostatic descriptor for hydrogen bond donors that can be used in the same way that V_{min} is used to characterize acceptors. This can be achieved by posing the question as to how MEP should be calculated in order for it to be optimally predictive of hydrogen bond acidity. The axis defined by the donor hydrogen and the atom to which it is bonded and distance, r, from the hydrogen nucleus on this axis provide a convenient reference frame (Figure 1) in which to address this question. A training set of 41 compounds with log K_α values spanning a range of 2.76 units, in which the donor hydrogen was bonded to either nitrogen (17 compounds) or oxygen (24 compounds), was selected. Measured values of log K_α and values of $V_\alpha(r)$ calculated for $r = 0.55$ Å at the B3LYP/6-31+G** level, using HF/6-31G* energy-minimized structures for the training set compounds (Chart 1), are presented in Table 1. The ability of eq 3 to fit the measured values of log K_α was quantified as a function of theoretical model, basis set and distance r from the hydrogen nucleus. The best-fitting linear models for each combination of theoretical method and basis set are presented in Table 2.

Comparable results are obtained for the three electronic structure models used in this study, and the ability of $V_\alpha(r)$ to fit the measured values of hydrogen bond acidity was found to be remarkably insensitive to the theoretical method used. The best fit to log K_α was achieved for $r = 0.55$ Å using $V_\alpha(r)$ values calculated at the B3LYP/6-31+G** level with HF/6-31G* molecular geometries (eq 3g), and this computationally inexpensive protocol was used exclusively for all subsequent calculations described in this article. The results presented in Table 2 show that $V_\alpha(r)$ is most effective

Table 1. Measured Values of Hydrogen Bond Acidity Used To Derive the Relationship between log K_α and $V_\alpha(r)$[f]

structure[a]	$V_\alpha(r)$[b]	log K_α[c]	pred log K_α[d]
1	0.3136	0.70	0.70
2	0.3106	0.64	0.57
3	0.3266	1.34	1.25
4	0.3338	1.45	1.55
5	0.3237	1.04	1.12
6	0.3169	0.48	0.84
7	0.3504	2.48	2.25
8	0.3537	2.47	2.38
9	0.3294	1.52	1.36
10[e]	0.3241	1.15	1.14
11	0.3123	0.90	0.64
12	0.3195	1.18	0.95
13	0.3242	0.73	1.15
14	0.3200	1.00	0.96
15	0.3234	0.95	1.11
16	0.3262	1.15	1.23
17	0.3381	1.20	1.73
18	0.3191	1.21	0.93
19	0.3186	1.11	0.91
20	0.3173	0.91	0.85
21	0.3150	0.78	0.76
22	0.3473	2.00	2.12
23	0.3609	2.83	2.69
24	0.3444	2.14	2.00
25	0.3432	1.75	1.95
26	0.3428	1.95	1.93
27	0.3416	1.85	1.88
28	0.3650	2.69	2.86
29	0.3421	1.89	1.90
30	0.3419	1.89	1.89
31	0.3536	2.50	2.38
32	0.3361	1.79	1.65
33	0.3572	2.56	2.53
34	0.3682	2.95	3.00
35	0.3495	2.44	2.21
36	0.3529	2.36	2.35
37	0.3401	2.18	1.81
38	0.3694	3.12	3.04
39	0.3479	2.04	2.14
40	0.3450	1.77	2.02
41	0.3479	2.07	2.14

[a] See Chart 1. [b] Calculated for r = 0.55 Å at the B3LYP/6-31+G** level using HF/6-31G* energy-minimized molecular geometry. [c] Measured hydrogen bond acidity; see refs 25, 44, and 45. [d] Predicted hydrogen bond acidity predicted using eq 3g (Table 2). [e] Two equivalent donor hydrogen atoms. [f] See Chart 1 for the structures of the training set compounds.

as a predictor of log K_α when calculated well within the van der Waals sphere of the donor hydrogen. It is possible to simulate the effect of using MEP calculated on a typical molecular surface to predict hydrogen bond acidity[38,39] by setting r to the van der Waals radius of hydrogen[57] (1.20 Å). The RMSE for eq 3h (0.434) is more than double than the figure for eq 3g (0.196).

The largest difference between measured and predicted values of log K_α in the training set was observed for **17**. The value predicted for this is 0.53 units too high, and one can speculate that this may reflect restriction of the rotation of the substituent chain by hydrogen bond formation. The log K_α values measured for phenol (**24**; pK_a= 10.0[58]) and benzoic acid (**41**; pK_a= 4.2[58]) differ by less than 0.1 units despite the almost 6 unit difference in pK_a, illustrating that proton acidity is not a good predictor of hydrogen bond acidity.[25] The apparent weakness of **41** as a hydrogen bond donor can be thought of as due to a repulsive secondary electrostatic interaction[59] between the carbonyl oxygen atom of the carboxylic acid and an acceptor of a hydrogen bond from the acidic hydrogen atom. Values of log K_α predicted for these compounds using eq 3g are both within 0.15 units of measured values, suggesting that this simple electrostatic model for hydrogen bonding adequately accounts for these secondary interactions.[59]

Measured values of log K_α were used to validate the predictive models for this quantity, and comparisons of these with values predicted using eq 3g are presented in Table 3. The approach to selecting the validation set has been to identify compounds which are unsuitable for inclusion in the training set or which might be expected, on the basis of particular structural features, to be challenging test cases. Compounds with nonequivalent acceptors have been shown[29,37] to be useful for validating predictive models for hydrogen bond basicity, and log K_α values for compounds (**46**, **47**) with nonequivalent donors were included in the validation set. Measured values of compounds for which there was uncertainty in tautomeric[60] (**42**, **43**, **44**) or conformational (**45**) preferences were also used. Two anilides (**50**, **51**) for which intramolecular hydrogen bonding has been proposed[45] to be an important determinant of hydrogen bond donor strength were also used for model validation. Modeling hydrogen bonding of these species is potentially challenging because the electronic effect of the ring is relayed to the hydroxyl group by an intramolecular hydrogen bond. In cases where the structure can exist in more than one low-energy conformational or tautomeric form, predicted values and electronic energy relative to that of the lowest energy form are presented. Values of K_α predicted for the individual donors in **46** and **47** were summed to generate figures for comparison with values measured for these compounds.

Predictions for the validation set compounds that were modeled in multiple forms were good. Values of log K_α predicted for different forms of a given molecule were generally similar, and in no case did the range in values span more than 0.21 units. It was for this reason that no attempt was made to use differences in energies calculated for the different forms to predict equilibrium populations of each form. The differences between the value of log K_α measured for a compound and the value predicted for a form of the compound range from −0.03 (**44**, form B) to 0.33 (**45**, form A). Agreement between the measured and predicted values is excellent for **47**, but the value of log K_α predicted for **46** is 0.58 units too high.

The results presented in Tables 1−3 show that $V_\alpha(r)$ is an effective predictor of hydrogen bond donor strength when calculated at an appropriate distance, r, from the hydrogen nucleus. It can be used to predict hydrogen bond donor strength when measured values are either unavailable or difficult to interpret, as is the case when nonequivalent donors are present in the molecule. Together with V_{min}, this property fits naturally into the quantitative structure property/activity relationship (QSPR/QSAR) paradigm and has the advantage of a more secure physical basis than some of the descriptors (e.g., topological;[61] PSA[31]) that are commonly used in these studies. In particular, $V_\alpha(r)$ provides a means to predict the contribution of hydrogen bond donors to ΔlogP just as that of hydrogen bond acceptors to this property can be modeled with V_{min}.[29]

Lactam Self-Association. Self-association in nonpolar solvents is of particular interest in interpretation of partition

Table 2. Fit of Eq 2 to Measured Hydrogen Bond Acidity

eq	r/Å[a]	theoretical model	geometry[b]	RMSE[c]	R-square	F	log $K_{\alpha,0}$	v(r)
3a	0.55	HF/6-31G*	HF/6-31G*	0.202	0.924	472	−11.886	41.161
3b	0.55	HF/6-31G**	HF/6-31G*	0.198	0.927	495	−12.023	40.484
3c	0.55	HF/6-31+G**	HF/6-31G*	0.197	0.928	500	−12.025	39.965
3d	0.55	HF/6-311G**	HF/6-31G*	0.207	0.920	449	−12.065	40.546
3e	0.50	B3LYP/6-31G*	HF/6-31G*	0.207	0.919	443	−14.493	41.857
3f	0.50	B3LYP/6-31G**	HF/6-31G*	0.210	0.917	434	−14.493	41.149
3g	0.55	B3LYP/6-31+G**	HF/6-31G*	0.196	0.928	505	−12.498	42.080
3h	1.20	B3LYP/6-31+G**	HF/6-31G*	0.434	0.647	71	−2.855	55.399
3i	0.45	B3LYP/6-31+G**	B3LYP/6-31G*	0.197	0.927	495	−17.104	38.620
3j	0.50	B3LYP/6-311G**	HF/6-31G*	0.209	0.918	440	−14.746	41.423
3k	0.55	MP2/6-31G*	HF/6-31G*	0.214	0.914	415	−12.800	43.928
3l	0.55	MP2/6-31G**	HF/6-31G*	0.219	0.910	396	−12.852	43.533
3m	0.55	MP2/6-31+G**	HF/6-31G*	0.214	0.914	416	−12.797	42.514
3n	0.55	MP2/6-311G**	HF/6-31G*	0.232	0.899	349	−13.186	44.389

[a] Value of r for which $V_\alpha(r)$ best fit measured log K_α for training set data. [b] Theoretical model and basis with which structures were energy-minimized. [c] Root mean square error.

Table 3. Comparison of Measured and Predicted Values of log K_α for Compounds in the Validation Set[I]

structure	log K_α[a]	donor[b]	energy (kJ mol^{-1})[c,d]	$V_\alpha(r)$[d,e]	pred log K_α[f]
42	2.71			0.3574	2.54
43	2.18	A	0.0	0.3421	1.90
		B	5.7	0.3424	1.91
44	1.99	A	0.0	0.3414	1.87
		B	0.9	0.3437	1.96
45	2.63	A	0.0	0.3673	2.96
		B	4.2	0.3623	2.75
46	1.52	total			2.10
		syn		0.3378	1.71
		anti		0.3414	1.87
47	1.1	total			1.16
		syn		0.3113	0.60
		anti		0.3214	1.05
48	1.58[g]			0.3375	1.70
49	3.55			0.3807	3.52
50	1.7[h]			0.3435	1.96
51	2.8[h]			0.3726	3.18

[a] From ref 25 unless unless otherwise specified. [b] Refers either to a specific form (tautomer; conformer) of a structure or to an individual hydrogen atom in structures with more than one donor. [c] The difference in energy between alternative forms of a compound. [d] B3LYP/6-31+G** electronic structure model using molecular geometry energy-minimized at the RHF/6-31G* level. [e] In atomic units (negative Hartree/electron). [f] Predicted from eq 3g. [g] Reference 44. [h] Reference 45. [I] See Figure 2 and Chart 1 for structures and specification of donors.

Table 4. Association Constants and MEP Values for Lactams as a Function of Ring Size

structure	K (IR)[a,b]	K (NMR)[a,c]	V_α(0.55 Å)[d]	V_{min} (*syn*)[d,e]	V_{min} (*anti*)[d,f]
52	51	61	0.3148	−0.0865	−0.0885
53	165	182	0.3076	−0.0927	−0.0958
54	131	139	0.2975	−0.0944	−0.0985
55	93	105	0.2972	−0.0947	−0.0982
56	77	99	0.2965	−0.0952	−0.0987

[a] Stability constant for dimer formation in units of mol^{-1} dm^3 from ref 62. [b] Measured by an infrared spectroscopic method. [c] Measured by the NMR spectroscopic method. [d] In atomic units (negative Hartree/electron). [e] Electrostatic potential minimum *syn* to donor hydrogen. [f] Electrostatic potential minimum *anti* to donor hydrogen.

Table 5. Calculated MEP for Hydrogen Bond Donors and Acceptors in Isosterically-Modified DNA Bases[b]

structure	template	$V_\alpha(r)$ (Do1)[a]	$V_\alpha(r)$ (Do2)[a]	V_{min} (Ac1)[a]	V_{min} (Ac2)[a]	V_{min} (N7)[a]
57	A	0.3048	0.3051	−0.0902		−0.0841
58	A	0.3122	0.3260	−0.0846		
59	A	0.2984	0.3117	−0.0946		
60	A	0.2878	0.2868	−0.1009		−0.0889
61	A	0.3200	0.3147	−0.0795		−0.0834
62	A	0.3291	0.3197	−0.0757		−0.0601
63	G	0.3355	0.3311	−0.0856	−0.1053	−0.1094
64	G	0.3399	0.3362	−0.0790	−0.0878	
65	G	0.3289	0.3255	−0.0893	−0.0984	
66	G	0.3240	0.3180	−0.0959	−0.1162	−0.1148
67	G	0.3392	0.3421	−0.0731	−0.0940	−0.1087
68	G	0.3411	0.3478	Not found	−0.0907	−0.1007

[a] In atomic units (negative Hartree/electron). [b] See Chart 2 and Figure 4 for structures and specification of hydrogen bond donors and acceptors.

coefficients since it can mask solute polarity.[29] The association constants (Table 4) measured in carbon tetrachloride for lactams ranging in ring size from 4 to 8 show the interesting feature of a maximum value for the 5-membered ring.[62] Values of $V_\alpha(r)$ and V_{min} calculated for these lactams suggest that donor strength decreases with ring size while acceptor strength increases. This trend may reflect differences in bond angles or the greater inductive effects of donor and acceptor on each other in smaller rings. The maximum in self-association constant observed for the five-membered lactam suggests that neither donor nor acceptor strength predominates in determining this property for these structures. The link between these calculations and molecular design is that they show how a composite molecular recognition feature, such as an adjacent donor and acceptor, can be dissected into individual elements of molecular recognition in a meaningful manner.

Isosterically Modified DNA Bases. Values of $V_\alpha(r)$ and V_{min} were calculated for isosterically modified purine nucleotide bases to illustrate how these properties might be used in hypothesis-driven design. The bases were modeled as their 9-methyl analogs and analysis focused on the Watson−Crick[7] donors and acceptors (Table 5; Figure 4; Chart 2). Structural modifications of the parent bases that led to an increase in $V_\alpha(r)$ typically resulted in V_{min} becoming

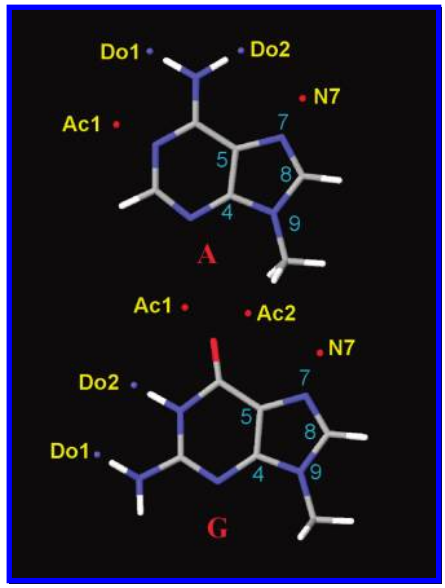


Figure 4. Electrostatic potential minima (red) and points (blue) at which $V_\alpha(r)$ was calculated to quantify effects of isosteric modification on DNA bases (see Table 5). The positions of the minima associated with the oxygen atom of 9-methylguanine reflect interactions with the N1 hydrogen atom and N7 lone pair.

less negative. These properties are highly correlated (ρ = 0.997) for the Watson−Crick donor (Do1) and acceptor (Ac1) in adenine isosteres. A principal component analysis[63] of the electrostatic potential values associated with Do1, Do2, and Ac1 (Figure 4) of the guanine isosteres showed that the first principal component accounts for 97% of the variance. In no case did isosteric modification of either adenine or guanine result in a simultaneous increase or decrease in the magnitudes of the $V_\alpha(r)$ and V_{min} values associated with Watson−Crick donors and acceptors. This observation suggests that these isosteric modifications can be described as either electron-withdrawing or electron-releasing with respect to the Watson−Crick hydrogen bond donors and acceptors. The relevance of these calculations to molecular design is that they allow the degree of electron withdrawal or release to be quantified in the context of the molecular system of interest rather than in a generic model system such as the ionization of benzoic acids used to define Hammett constants.[64]

Calculations for guanine isosteres show how $V_\alpha(r)$ and V_{min} can be used to provide a rationale for the effects of structural changes on duplex stability. Replacement of N7 of guanosine with carbon has been reported[65] to lower melting temperature by 9 °C and duplex stability by 3 kJ mol^{-1} in each of the hexamers d(CGCGCG) and d(GCGCGC). This isosteric modification of the DNA can be defined compactly in terms of the model structures as [**63**→**65**], and this notation will be used to specify the other modifications of DNA discussed in this article. The corresponding figures for [**63**→**64**] are a 13−16 °C increase in melting point and a 10−13 kJ mol^{-1} increase in duplex stability.[65] In larger (10−12) oligonucleotides, a single [**63**→**65**] modification was observed to decrease duplex stability by 0.8 kcal mol^{-1} to 4.8 kcal mol^{-1} depending on ionic strength.[66] The decrease in the values of $V_\alpha(r)$ and V_{min} resulting from [**63**→**65**] shows that this modification of DNA which destabilizes duplex formation is electron-releasing. Analogously, [**63**→**64**], which stabilizes duplex formation, is shown to be electron-withdrawing with respect to the Watson−Crick donors and acceptor of guanosine.

Guanine contributes two hydrogen bond donors and one acceptor to the GC base pair, and one can pose the hypothesis that donor strength of guanine is more important than its acceptor strength in determining the contribution of the base pair to duplex stability. Brought together, the changes in duplex stability observed for [**63**→**64**] and [**63**→**65**] and calculated values of $V_\alpha(r)$ and V_{min} support this hypothesis. These results are also consistent with the view that the hydrogen bonding properties of guanine do not appear to have been optimized for duplex stabilization even if they have been for physiological function.

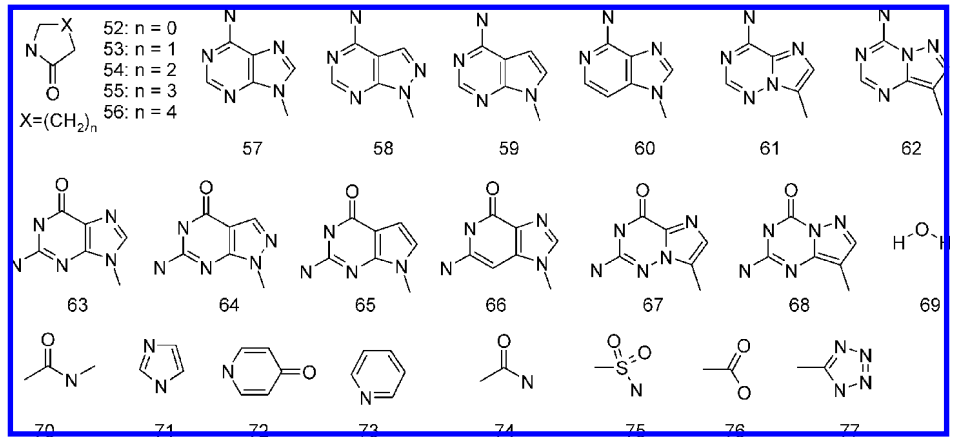
The nitrogen atom at position 9 can be moved to either ring-fusion position although these isosteric modifications are less frequently encountered in the nucleic acid literature. These are of interest because the hydrogen bonding geometry of the parent bases is retained. Calculated values of $V_\alpha(r)$ and V_{min} suggest that these isosteric modifications should strengthen donors and weaken acceptors associated with Watson−Crick base pairing. The effects for both adenine and guanine were more pronounced for moving N9 to position 5 than position 4. On this basis, it is expected that the [**63**→**67**] and [**63**→**68**] modifications will lead to an increase in duplex stability and that a larger increase will be observed for the latter.

Structures **62** and **68** are of particular interest because the V_{min} values associated with N7 in these structures are significantly less negative than the corresponding values for the parent heterocycles. The [**57**→**62**] and [**63**→**68**] modifications are expected to weaken interactions of the N7 acceptor relative to the parent heterocycles. The difficulties presented by GC-rich regions in using the polymerase chain reaction[67] (PCR) are believed to be due both to the trifurcated nature of the Watson−Crick interaction and the ability of N7 to accept hydrogen bonds, and 7-deazaguanosine has been used with some success as a replacement for guanosine in this procedure.[68] The [**63**→**68**] modification provides a means to dissect these two factors since it is likely to increase the strength of the Watson−Crick interaction while weakening the interactions of N7.

The N7 hydrogen bond acceptor can also function as a nucleophile. The most negative V_{min} value found for nitrogen atoms in the parent bases is that associated with N7 of **63** which is consistent with N7 of guanine being the predominant site for alkylation of DNA.[69] The V_{min} values calculated for N7 in the isosterically modified DNA bases suggest that [**57**→**62**] and [**63**→**68**] would reduce N7 alkylation to the greatest extent without totally eliminating it.

Effect of Complex Formation on Hydrogen Bond Donor Strength. Water molecules in contact with proteins can be viewed as molecular recognition elements in their own right, and there is considerable interest in their properties.[70,71] Hydrogen bond formation perturbs unused donors and acceptors in the interacting molecules which is particularly relevant to development and parametrization of polarizable force fields.[72,73] Figure 5 shows schematically how the effects of complex formation on other hydrogen bond donors and acceptors in the system may be quantified. Measuring stability constants for ternary complexes is technically demanding because pairwise interactions between

Chart 2. Structures for Studies of Effects of Isosteric Modification of DNA Bases and Complex Formation on Hydrogen Bond Acidity



three species must be considered and BSSE correction is also more complicated for these systems. It has been shown previously that V_{min} can provide insight into the effects of complex formation on hydrogen bond basicity,[29] suggesting that $V_\alpha(r)$ might be used in an analogous manner for hydrogen bond acidity.

Hydrogen bond formation results in intermolecular charge transfer (CT) from the acceptor to the donor. Complex formation is therefore expected to strengthen any donors in the acceptor component of the complex while weakening any unused donors in the donor component.[74] In this study, values of $V_\alpha(r)$ were compared for intermolecular complexes and their parent molecules in order to quantify the effect of complex formation on hydrogen bond donor strength. Each complex included water, either as the hydrogen bond donor or acceptor component, and the effect of complex formation was quantified by the resulting difference, $\Delta V_\alpha(r)$, in $V_\alpha(r)$. The calculations are grouped according to whether the hydrogen bond donor being probed was in the acceptor (Table 6) or donor (Table 7) component of the complex. The notation **69.71** is used to specify the complex in which **71** accepts a hydrogen bond from **69** and complexes in which the acceptor component also functions as a donor are specified using parentheses as **(69.)$_3$** or **74.(69.)** (Figure 6).

The results presented in Table 6 show that the $\Delta V_\alpha(r)$ value calculated for water dimer, **69.69**, is greater than those observed for the complexes **69.70a**, **69.70b**, **69.71**, or **69.72**. The small distance between donor and probe hydrogen atoms in **69.69** provides an obvious rationale for this observation. Eq 3g can be used to show that accepting a hydrogen bond from another water molecule is likely to increase donor strength on the log K_α scale from an 'alcohol-like' value of 1.2 to a 'phenol-like' 2.0. Prediction of log K_α from $V_\alpha(r)$ values of hydrogen-bonded complexes is useful because it allows the effects of complex formation to be compared with those resulting from structural variation. As might be expected, the strength of the donor influences the magnitude of the $\Delta V_\alpha(r)$ values observed for the hydrogen atoms in the acceptor component of the complex. The donor of **72** (predicted log $K_\alpha = 2.21$) is particularly strong, and the value of $V_\alpha(r)$ calculated for **72.69** (0.3641) suggests that accepting a hydrogen bond from **72** will increase the log K_α value of

Table 6. Effect of Complex Formation on Donors in Acceptor Components of Complexes[d]

complex	donor	acceptor	$V_\alpha(r)^{a,b}$	$V_\alpha(r)$ (complex)^{a,c}	$\Delta V_\alpha(r)^a$
69.69	**69**	**69**	0.3259	0.3439	+0.0180
(69.)$_3$	**69**	**69**	0.3259	0.3275	+0.0016
				0.3255	−0.0004
				0.3246	−0.0013
69.70a	**69**	**70**	0.3160	0.3248	+0.0088
69.70b	**69**	**70**	0.3160	0.3245	+0.0085
69.71	**69**	**71**	0.3386	0.3453	+0.0077
69.72	**69**	**72**	0.3496	0.3555	+0.0059
70.69	**70**	**69**	0.3259	0.3516	+0.0257
71.69	**71**	**69**	0.3259	0.3558	+0.0299
72.69	**72**	**69**	0.3259	0.3641	+0.0382

[a] In atomic units (negative Hartree/electron). [b] Value in an uncomplexed molecule. [c] Value in a complex. [d] See Figure 6 and Chart 2 for specification of structures.

Table 7. Effect of Complex Formation on Strength of Unused Hydrogen Bond Donors in Donor Components of Complexes[d]

complex	donor	acceptor	$V_\alpha(r)^{a,b}$	$V_\alpha(r)$ (complexed)^{a,c}	$\Delta V_\alpha(r)^a$
69.69	**69**	**69**	0.3259	0.3051	−0.0208
69.71	**69**	**71**	0.3259	0.2982	−0.0277
69.73	**69**	**73**	0.3259	0.3008	−0.0251
69.72	**69**	**72**	0.3259	0.2914	−0.0345
74.(69.)	**74**	**69**	0.3209	0.3222	+0.0013
74.69	**74**	**69**	0.3081	0.2924	−0.0157
75.(69.)	**75**	**69**	0.3243	0.3242	−0.0001

[a] In atomic units (negative Hartree/electron). [b] Value in an uncomplexed molecule. [c] Value in a complex. [d] See Figure 6 and Chart 2 for specification of structures.

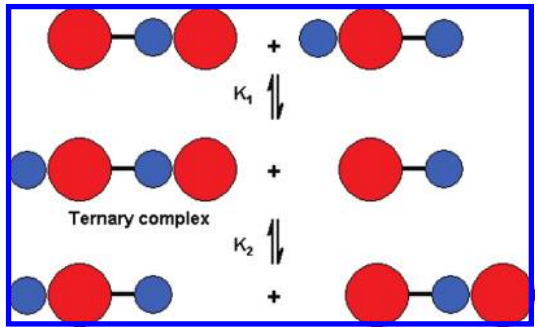


Figure 5. Effects of complex formation on hydrogen bond strength can be quantified by the extent to which the equilibrium constants K_1 and K_2 differ from unity. Hydrogen bond donors and acceptors are represented as blue and red circles, respectively.

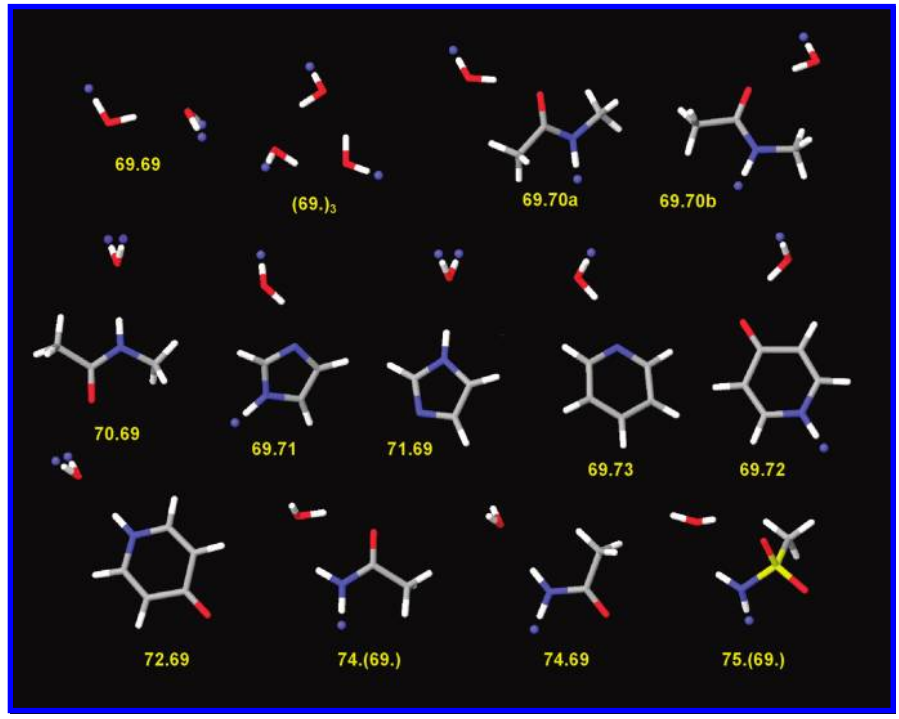


Figure 6. Structures of hydrogen bonded complexes showing points (colored blue) at which MEP was calculated.

water to 2.82 which is similar to the figure of 2.80 measured[25] for 4-trifluoromethylphenol.

Table 7 summarizes how donating a hydrogen bond weakens the other hydrogen bond donors in a molecule. The magnitude of the effect increases with the hydrogen bond basicity of the acceptor. This is illustrated by the $\Delta V_\alpha(r)$ values calculated for the water dimer (−0.0208) and the complex of water with the very strong acceptor[29] **72** (−0.0345). The negative value of $\Delta V_\alpha(r)$ shown in Table 7 for complex **69.73** can be used to provide a rationale for the observation that the 1−2 complex of water with pyridine is over 10-fold less stable than the 1−1 complex.[75] The $\Delta V_\alpha(r)$ values for **69.73** (−0.0251) and **69.71** (−0.0277) suggest that this effect would be larger for imidazole than for pyridine. The $\Delta V_\alpha(r)$ value of −0.0157 observed for **74.69** suggests that accepting a hydrogen bond from the hydrogen atom in **74** that is *anti* with respect to the carbonyl oxygen will lead to a significant reduction in the hydrogen bond acidity of the *syn* donor. The $V_\alpha(r)$ value for the *anti* donor of **74** (0.3209) is significantly greater than that of the *syn* donor (0.3081), reflecting secondary electrostatic interactions between the latter hydrogen atom and the adjacent lone pair of the carbonyl oxygen.[59] These differences in hydrogen bond acidity will be amplified in aqueous solution because both donors can simultaneously form hydrogen bonds, which provides some justification for counting an amino group as a single donor when applying the rule of 5.[30]

Cyclic water trimer, **(69.)$_3$**, shows hydrogen bonding characteristics that are very different from those of **69.69** (Table 6). The three hydrogen atoms of water trimer not directly involved in hydrogen bonding experience the opposing effects of water functioning both as donor and acceptor, and $V_\alpha(r)$ values are all very similar to that observed for the isolated water molecule. This result supports the view, expressed by Latimer and Rodebush in 1920, that water 'shows tendencies both to add and give up hydrogen, which are nearly balanced'.[1] Analogously, the $\Delta V_\alpha(r)$ values

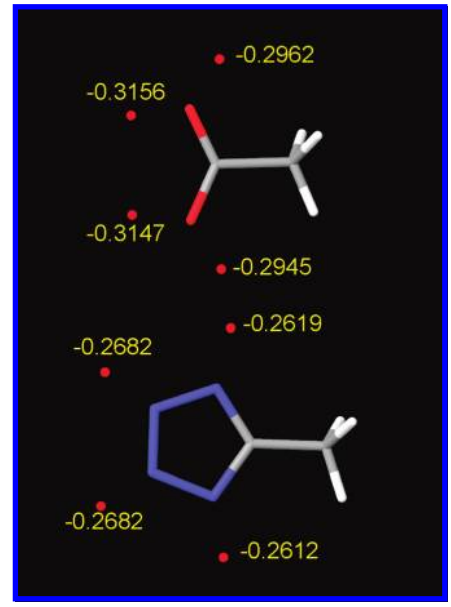


Figure 7. Electrostatic potential minima for anionic forms illustrate the bioisoteric relationship between the anionic form of carboxylic acids (top) and corresponding tetrazoles (bottom).

observed for amide and sulfonamide donor hydrogen atoms in complexes **74.(69.)** and **75.(69.)** are both very small in magnitude.

Bioisosteric Relationship between Carboxylic Acids and Tetrazoles. Bioisosteres are molecular recognition features that have similar properties with respect to recognition by protein molecules, and biososteric relationships are exploited in medicinal chemistry to increase potency and manipulate physicochemical properties.[76−78] The relationship between tetrazoles[79] and the corresponding carboxylic acids provides an excellent and arguably prototypical example of the phenomenon. Tetrazoles and carboxylic acids are deprotonated under normal physiological conditions, and it has been shown that replacement of a carboxylic acid with

tetrazole is likely to lead to increased binding to plasma protein.[80] The four MEP minima observed (see Figure 6) for each of the anionic forms of acetic acid and 5-methyltetrazole show the similarity between these bioisosteres in a way that is simply not possible using atom-centered charges. These calculations also reveal subtle differences between **76** and **77** in that the V_{min} values for **77** are all mutually similar. In contrast the minima furthest from the linking point (modeled as methyl) in **76** are significantly more negative than the other two.

Implications for Computational Chemistry. The results presented in this study have wider implications for computational chemistry and molecular modeling. Molecular recognition can be viewed as mutual presentation of molecular surfaces which provides a rationale for using MEP on the molecular surface as a QSPR descriptor with which to model solvation.[39,40] The observation that the RMSE for eq 3h (r = 1.20 Å; van der Waals radius for hydrogen) is more than twice that for eq 3g (r = 0.55 Å) provides guidance as to how molecular surfaces might be constructed in order to best model hydrogen bond donors. However, locating points at which $V_\alpha(r)$ is calculated actually represents a computationally less demanding procedure than locating electrostatic potential maxima on molecular surfaces. The results of this study bring into question the value of molecular surfaces for predicting hydrogen bond acidity.

The molecular surface, or more accurately the dielectric boundary,[81] is an essential component of continuum solvent models, and it might be thought that the results of this study would also be relevant to these approaches to modeling solvation. However, other factors such as the outlying charge problem[82] need to be considered when setting radii for continuum solvent models. The widely adopted Generalized Born[83] and Poisson−Boltzmann[84] models are symmetrical[85] with respect to charge type and yield identical solvation energies for ions of opposite charge and equal radius. In water, donor and acceptor strength are comparable, and this issue can be addressed, to some extent, by manipulation of atomic radii. However, this is unlikely to represent a viable approach to modeling dipolar, aprotic solvents such as dimethyl sulfoxide and dimethylformamide, which are known to be strong hydrogen bond acceptors.[25] This line of reasoning can be extended to 1-octanol, the default solvent for measurement of lipophilicity,[86] which has only a single hydrogen bond donor, and it may prove more challenging than water to represent as a continuum on the basis that its donor and acceptor strengths are less comparable. It should be noted that the SM8[87] and Cosmo-RS[88] continuum models are not symmetrical with respect to charge type and might be expected to treat dipolar aprotic solvents in a physically meaningful manner.

The results of this study also have implications for deriving atomic charges. It is common to fit these to values of MEP calculated on a molecule-centered grid, the points of which typically lie outside the atomic van der Waals spheres.[34] This is done 'to avoid getting too close to nuclei where the electrostatic potential is always positive'.[89] However, the electrostatic potential minima used to model acceptors and the points at which $V_\alpha(r)$ is most predictive of hydrogen bond donor strength are located within the van der Waals spheres of their associated atoms. This ensures that MEP is not sampled at the points where this property is most predictive of hydrogen bond acidity. An analogous argument can be made for calculating MEP within the van der Waals spheres of donor hydrogen atoms in CoMFA[35] studies in which MEP is calculated using quantum mechanical models.[37]

The calculations for hydrogen-bonded complexes are relevant to development of polarizable force fields.[72,73] An earlier study[29] suggested that polarization effects are likely to be significant for hydrogen bonds involving carbonyl and hydroxyl groups. The current study suggests that polarization is likely to be important when nitrogen is bound to two or more hydrogen atoms even in situations where the nitrogen atom does not function as a hydrogen bond acceptor (e.g., primary amides). Information of this nature might be used in force field development to restrict treatment of polarization to those atom types for which the effects are likely to be most significant.

CONCLUSIONS

This article shows that MEP, calculated at a relatively low level of theory, is an effective predictor of hydrogen bond acidity. Calculated values of $V_\alpha(r)$, which has been introduced in this study, and V_{min} provide insight into a range of hydrogen bonding phenomena. For example, the extent of electron withdrawal and release can be quantified in a relevant structural context, and effects of complex formation and structural modification can be brought onto a common scale. Implications of these results for a number for areas of computational chemistry are discussed, including use of molecular surfaces to predict hydrogen bond acidity, fitting atomic charges to MEP, and development of polarizable force fields.

Abbreviations: B3LYP, Becke three-parameter exchange functional with Lee−Yang−Parr correlation functional; BSSE, basis set superposition error; CoMFA, comparative molecular field analysis; COSMO-RS, conductor-like screening model for real solvents; CT, charge transfer; HF, Hartree−Fock; MEP, molecular electrostatic potential; MP2, second order Møller−Plesset perturbation theory; PCR, polymerase chain reaction; PSA, polar surface area; QSAR, quantitative structure activity relationship; QSPR, quantitative structure property relationship; RMSE, root-mean-square error; SM8, solvation model; $V_\alpha(r)$, electrostatic potential at distance, r, from hydrogen on axis defined by hydrogen and atom to which it is bonded; V_{min}, minimized electrostatic potential; ΔlogP, difference between logarithms of 1-octanol/water and hydrocarbon/water partition coefficients.

ACKNOWLEDGMENT

The author thanks Amy Ginsburg, Karen Kirkham, and Andrew Maynard for providing insight into the molecular recognition process and Andrew Leach and Claire Gavaghan for helpful and constructive comments on the manuscript.

Supporting Information Available: Molecular geometries energy-minimized at the HF/6-31G* level of theory and the full Gaussian 03 citation. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Latimer, W. M.; Rodebush, W. H. Polarity and ionization from the standpoint of the Lewis theory of valence. *J. Am. Chem. Soc.* **1920**, *42*, 1419–1433.

(2) Pimental, G. C.; McClellan, A. L. *The Hydrogen Bond*; Freeman: San Francisco, 1960.
(3) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, 1997.
(4) Etter, M. C. Encoding and decoding hydrogen-bond patterns of organic compounds. *Acc. Chem. Res.* **1990**, *23*, 120–126.
(5) Bernstein, J.; Davis, R. E.; Shimoni, L.; Chang, N.-L. Patterns in hydrogen bonding: functionality and graph set analysis in crystals. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1555–1573.
(6) Karamertzanis, P. G.; Day, G. M.; Welch, G. W. A.; Kendrick, J.; Leusen, F. J. J.; Neumann, M. A.; Price, S. L. Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs. *J. Chem. Phys.* **2008**, *128*, 244708/1–244708/17.
(7) Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature (London)* **1953**, *171*, 737–738.
(8) Franklin, R. E.; Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature (London)* **1953**, *171*, 740–741.
(9) Pauling, L.; Corey, R. B.; Branson, H. R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205–11.
(10) Pauling, L.; Corey, R. B. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 729–40.
(11) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97–179.
(12) Powers, E. T.; Deechongkit, S.; Kelly, J. W. Backbone-backbone H-bonds make context-dependent contributions to protein folding kinetics and thermodynamics: lessons from amide-to-ester mutations. *Adv. Protein Chem.* **2006**, *72*, 39–78.
(13) Baldwin, R. L. Energetics of protein folding. *J. Mol. Biol.* **2007**, *371*, 283–301.
(14) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys. Biomol. Struct.* **2008**, *37*, 289–316.
(15) Wells, T. N. C.; Fersht, A. R. Hydrogen bonding in enzymatic catalysis analyzed by protein engineering. *Nature (London)* **1985**, *316*, 566–657.
(16) Schowen, K. B.; Limbach, H.-H.; Denisov, G. S.; Schowen, R. L. *Biochim. Biophys. Acta* **2000**, *1458*, 43–62.
(17) Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, *14*, 1–63.
(18) Tanford, C. *The Hydrophobic Effect*, 2nd ed.; John Wiley: New York, 1980.
(19) Tanford, C. How protein chemists learned about the hydrophobic factor. *Protein Sci.* **1997**, *6*, 1358–1366.
(20) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. A view of the hydrophobic effect. *J. Phys. Chem. B* **2002**, *106*, 521–533.
(21) Taft, R. W.; Gurka, D.; Joris, L.; Schleyer, P. v R.; Rakshys, J. W. Studies of hydrogen-bonded complex formation with p-fluorophenol. V. Linear free energy relationships with OH reference acids. *J. Am. Chem. Soc.* **1969**, *91*, 4801–8.
(22) Kamlet, M. J.; Taft, R. W. The solvatochromic comparison method. I. The β-scale of solvent hydrogen-bond acceptor (HBA) basicities. *J. Am. Chem. Soc.* **1976**, *98*, 377–383.
(23) Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **1993**, *22*, 73–83.
(24) Laurence, C.; Berthelot, M. Observations on the strength of hydrogen bonding. *Perspect. Drug Discovery Des.* **2000**, *18*, 39–60.
(25) Abraham, M. H.; Duce, P. D.; Prior, D. V.; Barratt, D. G.; Morris, J. J.; Taylor, P. J. Hydrogen bonding. Part 9. Solute proton-donor and proton-acceptor scales for use in drug design. *J. Chem. Soc., Perkin Trans. 2* **1989**, 1355–1375.
(26) Seiler, P. Interconversion of lipophilicites from hydrocarbon/water systems into the octanol/water system. *Eur. J. Med. Chem.* **1974**, *9*, 473–479.
(27) Young, R. C.; Mitchell, R. C.; Brown, T. H.; Ganellin, C. R.; Griffiths, Jones, R. M.; Rana, K.K.; Saunders, D.; Smith, I. R.; Sore, N.; Wilks, T. J. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists. *J. Med. Chem.* **1988**, *31*, 656–671.
(28) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the ΔlogP parameter of Seiler. *J. Pharm. Sci.* **1994**, *83*, 1085–1100.
(29) Toulmin, A.; Wood, J. M.; Kenny, P. W. Toward Prediction of Alkane/Water Partition Coefficients. *J. Med. Chem.* **2008**, *51*, 3720–30.
(30) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.
(31) Palm, K.; Luthman, K.; Ungell, A.-L.; Strandlund, G.; Artursson, P. Correlation of Drug Absorption with Molecular Surface Properties. *J. Pharm. Sci.* **1996**, *85*, 32–39.
(32) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(33) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening - an overview. *Drug Discovery Today* **1998**, *3*, 160–178.
(34) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–80.
(35) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(36) Murray, J. S.; Ranganathan, S.; Politzer, P. Correlations between the solvent hydrogen bond acceptor parameter β and the calculated molecular electrostatic potential. *J. Org. Chem.* **1991**, *56*, 3734–3739.
(37) Kenny, P. W. Prediction of hydrogen bond basicity from computed molecular electrostatic properties: implications for comparative molecular field analysis. *J. Chem. Soc., Perkin Trans. 2* **1994**, 199–202.
(38) Murray, J. S.; Politzer, P. Correlations between the solvent hydrogen-bond-donating parameter α and the calculated molecular surface electrostatic potential. *J. Org. Chem.* **1991**, *56*, 6715–17.
(39) Hunter, C. A. Quantifying intermolecular interactions: Guidelines for the molecular recognition toolbox. *Angew. Chem., Int. Ed.* **2004**, *43*, 5310–5324.
(40) Lamarche, O.; Platts, J. A. Complementary nature of hydrogen bond basicity and acidity scales from electrostatic and atoms in molecules properties. *Phys. Chem. Chem. Phys.* **2003**, *5*, 677–684.
(41) Boys, S. F.; Bernadi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
(42) Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers. *J. Chem. Phys.* **1996**, *105*, 11024–11031.
(43) Kangas, E.; Tidor, B. Electrostatic Complementarity at Ligand Binding Sites: Application to Chorismate Mutase. *J. Phys. Chem.* **2001**, *105*, 880–888.
(44) Abraham, M. H.; Duce, P. P.; Morris, J. J.; Taylor, P. J. Hydrogen bonding. Part 2. Equilibrium constants and enthalpies of complexation for 72 monomeric hydrogen-bond acids with N-methylpyrrolidinone in 1,1,1-trichloroethane. *J. Chem. Soc., Faraday Trans. 1* **1987**, *83*, 2867–2881.
(45) Morris, J. J.; Hughes, L. R.; Glen, A. T.; Taylor, P. J. Non-steroidal antiandrogens. Design of novel compounds based on an infrared study of the dominant conformation and hydrogen-bonding properties of a series of anilide antiandrogens. *J. Med. Chem.* **1991**, *34*, 447–455.
(46) *Gaussian 03, Revision B.05*; Gaussian Inc.: 340 Quinnipiac St. Bldg 40, Wallingford, CT 06492, USA. http://www.gaussian.com (accessed Mar 11, 2009).
(47) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory*; Dover: Mineola, 1996.
(48) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. IX. Extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
(49) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods. 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.
(50) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(51) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *3*, 785–789.
(52) Møller, C.; Plesset, M. S. Note on the approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618–622.
(53) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A direct MP2 gradient method. *Chem. Phys. Lett.* **1990**, *166*, 275–280.
(54) Spitznagel, G. W.; Clark, T.; Chandrasekhar, J.; Schleyer, P. v R. Stabilization of methyl anions by first-row substituents. The superiority of diffuse function-augmented basis sets for anion calculations. *J. Comput. Chem.* **1982**, *3*, 363–71.
(55) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–4.
(56) *JMP Version 6.0.0*; SAS Institute: SAS Campus Drive, Building S, Cary, NC 27513, USA. http://www.jmp.com (accessed Mar 11, 2009).
(57) Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
(58) Serjeant, E. P.; Dempsey, B. *Ionisation Constants of Organic Acids in Aqueous Solution*; Pergamon: Oxford, 1979.

(59) Jorgensen, W. L.; Pranata, J. Importance of secondary interactions in triply hydrogen bonded complexes: guanine-cytosine vs uracil-2,6-diaminopyridine. *J. Am. Chem. Soc.* **1990**, *112*, 2008–10.
(60) Elguero, J.; Marzin, C.; Katritzky, A. R.; Linda, P. *The Tautomerism of Heterocycles*; Academic Press: New York, 1976.
(61) Kier, L. B.; Hall, L. H. The prediction of ADMET properties using structure information representations. *Chem. Biodiversity* **2005**, *2*, 1428–1437.
(62) Prokopenko, N. A.; Bethea, I. A.; Clemens, C. J.; Klimek, A.; Wargo, K.; Spivey, C.; Waziri, K.; Grushow, A. The effect of structure on hydrogen bonding: Hydrogen bonded lactam dimers in CCl_4. *Phys. Chem. Chem. Phys.* **2002**, *4*, 490–495.
(63) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; Prentice Hall International: Englewood Cliffs, 1988.
(64) Hammett, L. P. Effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
(65) Seela, F.; Driller, H. Alternating d(G-C)3 and d(C-G)3 hexanucleotides containing 7-deaza-2′-deoxyguanosine or 8-aza-7-deaza-2′-deoxyguanosine in place of dG. *Nucleic Acids Res.* **1989**, *17*, 901–910.
(66) Ganguly, M.; Wang, F.; Kaushik, M.; Stone, M. P.; Marky, L. A.; Gold, B. A study of 7-deaza-2′-deoxyguanosine-2′-deoxycytidine base pairing in DNA. *Nucleic Acids Res.* **2007**, *35*, 6181–6195.
(67) Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. Primer-directed enzymic amplification of DNA with a thermostable DNA polymerase. *Science* **1988**, *239*, 487–91.
(68) Bachmann, H. S.; Siffert, W.; Frey, U. H. Successful amplification of extremely GC-rich promoter regions using a novel 'slowdown PCR' technique. *Pharmacogenetics* **2003**, *13*, 759–766.
(69) Gold, B.; Marky, L. M.; Stone, M. P.; Williams, L. D. A review of the role of the sequence-dependent electrostatic landscape in DNA alkylation patterns. *Chem. Res. Toxicol.* **2006**, *19*, 1402–1414.
(70) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling Water Molecules in Protein-Ligand Docking Using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.
(71) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 808–813.
(72) Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. **A** Polarizable Force Field and Continuum Solvation Methodology for Modeling of Protein-Ligand Interactions. *J. Chem. Theory Comput.* **2005**, *1*, 694–715.
(73) Xie, W.; Pu, J.; MacKerell, A. D.; Gao, J. Development of a Polarizable Intermolecular Potential Function (PIPF) for Liquid Amides and Alkanes. *J. Chem. Theory Comput.* **2007**, *3*, 1878–1889.
(74) Ceccarelli, C.; Jeffrey, G. A.; Taylor, R. A survey of O-H. .O hydrogen bond geometries determined by neutron diffraction. *J. Mol. Struct. Phys. Chem.* **1981**, *70*, 255–271.
(75) Lomas, J. S.; Maurel, F. Water and alcohol(s): what's the difference? A proton NMR and DFT study of hetero-association with pyridine. *J. Phys. Org. Chem.* **2008**, *21*, 464–471.
(76) Thornber, C. W. Isosterism and molecular modification in drug design. *Chem. Soc. Rev.* **1979**, *8*, 563–80.
(77) Patani, G. A.; LaVoie, E. J. Bioisosterism: A rational approach in drug design. *Chem. Rev.* **1996**, *96*, 3147–3176.
(78) Sheridan, R. P. The most common chemical replacements in drug-like compounds. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 103–108.
(79) Herr, R. J. 5-Substituted 1H-tetrazoles as carboxylic acid isosteres: medicinal chemistry and synthetic methods. *Bioorg. Med. Chem.* **2002**, *10*, 3379–3393.
(80) Birch, A. M.; Kenny, P. W.; Simpson, I.; Whittamore, P. R. O. Matched molecular pair analysis of activity and properties of glycogen phosphorylase inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 850–853.
(81) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608–640.
(82) Klamt, A.; Jonas, V. Treatment of the outlying charge in continuum solvation models. *J. Chem. Phys.* **1996**, *105*, 9972–9981.
(83) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **2000**, *90*, 6127–6129.
(84) Gilson, M. K.; Honig, B. H. Energetics of charge-charge interactions in proteins. *Proteins: Struct., Funct., Genet.* **1988**, *3*, 32–52.
(85) Mobley, D. L.; Barber, A. E.; Fennell, C. J.; Dill, K. A. Charge asymmetries in hydration of polar solutes. *J. Phys. Chem. B* **2008**, *112*, 2405–2414.
(86) Fujita, T.; Iwasa, J.; Hansch, C. A new substituent constant, π, derived from partition coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180.
(87) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033.
(88) Klamt, A.; Eckert, F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.* **2000**, *172*, 43–72.
(89) Cox, S. R.; Williams, D. E. Representation of the molecular electrostatic potential by a net atomic charge model. *J. Comput. Chem.* **1981**, *2*, 304–323.

CI9000234