Rank Correlation of AM1 and PM3 Derived Molecular Electrostatic Potentials (RACEL) with Hammett σ_p-Parameters

Howard B. Broughton,†[a] Stuart M. Green[b] and Henry S. Rzepa[b]

[a] *James Black Foundation, 68 Half Moon Lane, Dulwich, London SE24 9JE, UK*
[b] *Department of Chemistry, Imperial College of Science Technology and Medicine, London SW7 2AY, UK*

Spearman non-parametric correlation coefficients obtained by comparison of the rank order of Hammett σ_p-parameters and the value of AM1 and PM3 derived molecular electrostatic potentials calculated for a range of *para*-substituted benzoic acids and amines show a chemically significant three-dimensional spatial distribution; the correlation is best for PM3, is improved for carboxylic acids by the inclusion of a solvent water molecule and is substantially decreased if point-charge derived electrostatic potentials are employed.

Molecular electrostatic potentials (MEPs) have long been considered as one of the more useful tools in predicting chemical reactivity[1,2] and biological activity.[3] Recently, MEPs evaluated rigorously using the semi-empirical MNDO and AM1 Hamiltonians have been systematically evaluated[4] and applied to *e.g.* the derivation of atom centred point charges[5] and at the PM3 level to rationalising stereoelectronic effects in π-facial hydrogen bonding.[6] Correlations of MEPs with chemical or biological parameters are conventionally performed for a single value of the potential at spatial coordinates defined by a local minimum or maximum[2,4] in the function, or *via* the CoMFA technique[7] in which a partial least-squares parametric analysis is used to construct a linear relationship between a numerical representation of the electrostatic and steric fields of a molecule and some measured property. We present here a new and simpler technique (RACEL) which involves correlating the rank ordering of the value of the calculated electrostatic potential with *e.g.* Hammett σ_p-parameters[8] for a series of aromatic acids and amines, and for which the three-dimensional spatial distribution in the resulting correlation coefficient[9] is shown to have chemical significance.

MEPs at the MNDO, AM1 and PM3 level were calculated as a three-dimensional grid of values $V(x,y,z)$‡ for a series of *p*-substituted benzoic acids, the corresponding monohydrated species and the *p*-substituted anilines using the following substitutents; X = H, OH, OMe, CO_2H, NMe_2, Me, CF_3, NO_2, CN, F, Cl, Br and I. The choice of Hammett σ_p-parameters for the statistical correlation was based on the quality of the available experimental data[8] and the close theoretical relationship between electrostatic potentials and the equilibrium protonation reaction from which σ_p is derived.[4] The non-parametric technique of Spearman[9] was used to calculate the rank correlation coefficient $\mathbf{r}_s(x,y,z)$ between $V(x,y,z)$ for the range of molecules and the corresponding σ_p-parameters. Physically, a value for $\mathbf{r}_s(x,y,z) = 1$ signifies that the rank order of the value of the MEP at a common spatial point (x,y,z) for the range of compounds is exactly the same as the rank of the σ_p-parameters. Alternatively, $\mathbf{r}_s = -1$ signifies completely opposing ranks, and $\mathbf{r}_s = 0$ signifies no correlation. Graphical representation of the three dimensional iso-valued contours of $\mathbf{r}_s(x,y,z)$ allows identification of those regions of the molecules for which the MEP correlates most highly with the experimentally derived property, *i.e.* in this example the Hammett σ_p-parameter. The CoMFA technique[7] uses a more complex partial least-squares parametric analysis of functions derived from electrostatic and steric fields of a molecule and suffers from the limitations of parametric methods in demanding both the rank order and the absolute magnitude of the input data. CoMFA also requires a linear relationship between calculable and observed properties. While simple point-charge–point-charge interactions should obey the linear free energy paradigm, longer range interactions of the type revealed by full quantum mechanically derived MEPs are less likely to do so. Furthermore, the rank correlation technique should also be applicable to potentially non-linear properties, such as descriptors of frontier orbitals.

Our results for carboxylic acids show >24% of points with a correlation above $\mathbf{r}_s(x,y,z) = 0.90$ for PM3 and AM1 both with and without a solvating water molecule (Fig. 1). The MNDO method showed no correlation above the level $\mathbf{r}_s(x,y,z) = 0.75$. The region of space where high correlations can be expected is not limited to *e.g.* minima in the electrostatic potential[2,4] and >10% of the evaluated grid of

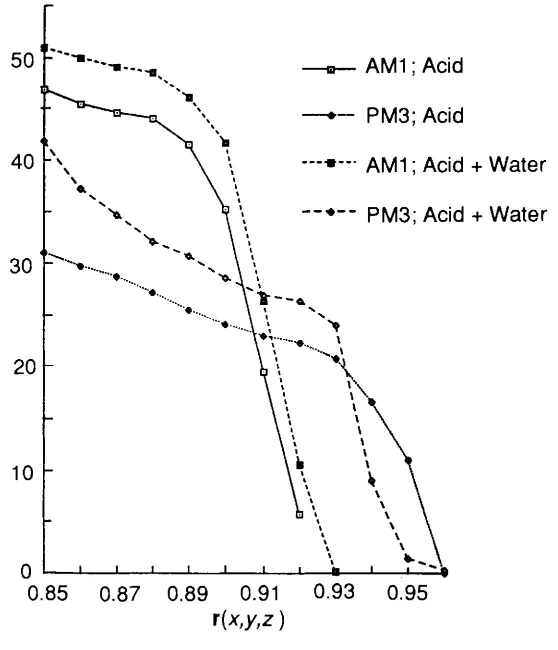

Fig. 1 Distribution of rank correlation coefficients between calculated values of the MEP and the Hammett σ_p-constants for substituted benzoic acids, expressed as a percentage having value $\geqslant \mathbf{r}(x,y,z)$

---

† *Present address*: Merck, Sharpe and Dohme Research Laboratories, Neuroscience Research Centre, Terlings Park, Eastwick Road, Harlow, Essex CM20 2QR.

‡ All calculations were performed on fully optimised molecular geometries using MOPAC (V6.0); J. J. P. Stewart, QCPE Program 455, locally modified such that the value of the electrostatic potential was evaluated throughout a regular 1 Å grid extending to 6 Å from the atom centres, ignoring points within the van der Waals radii. This grid definition was chosen as it yielded a computationally reasonable (*ca.* 3600) number of points, in which unfavourable steric interactions are implicitly excluded. Orientation of the molecules was kept consistent by ensuring the same first three atoms of the benzene ring (starting at the carbon bearing the CO_2 or NH_2 group) were used to define the internal coordinate system for each substituent. For anilines, one correlation was performed instead with the HNH atoms fixed, the sense of pyramidalisation of the nitrogen being the same for all molecules. For molecules in which there were two minimum energy conformations for the substituent (OH, OMe and CO_2H) MEPs were derived for both orientations of the substituent and both were included in the rank correlation analysis.

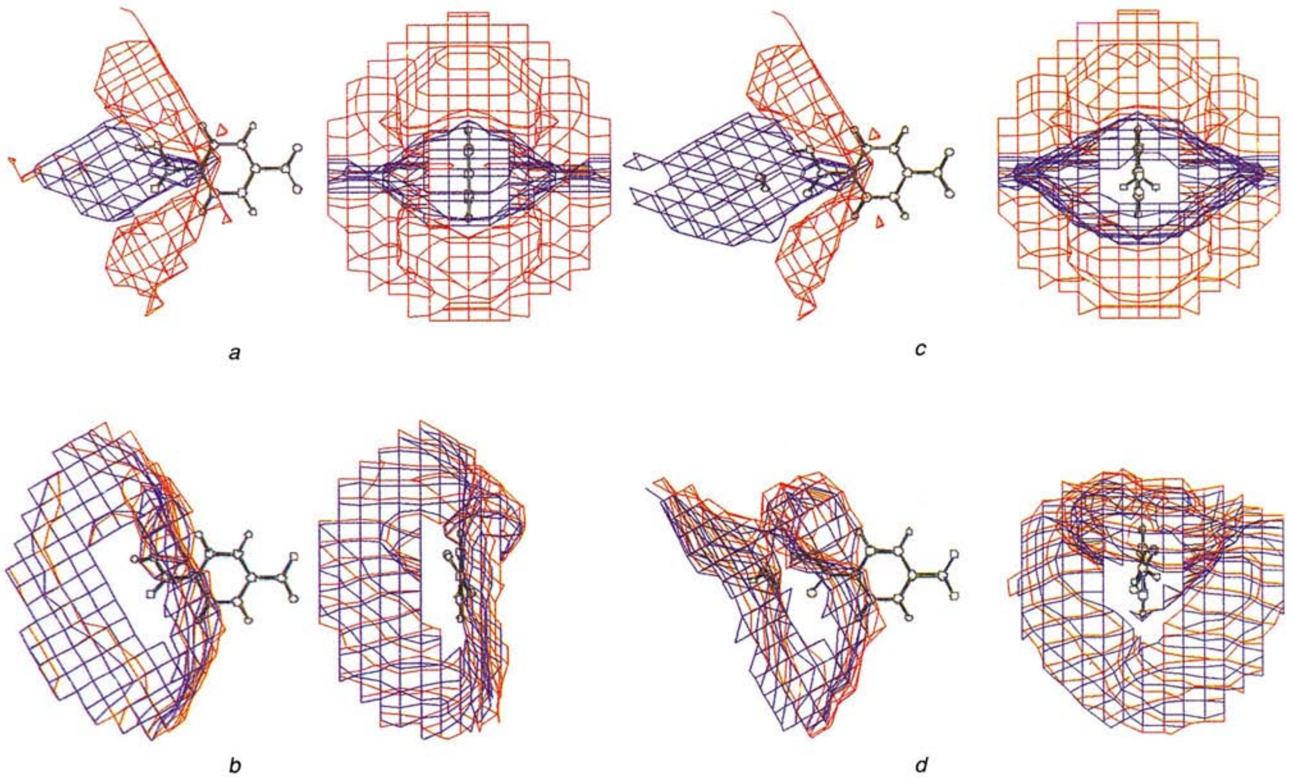

Fig. 2 Three dimensional distribution of values of $\mathbf{r}(x,y,z)$, contoured at $\mathbf{r} = 0.90$ (red) and $\mathbf{r} = 0.92$ (blue) for substituted benzoic acids, shown as two orthogonal views for (*a*) AM1 and (*b*) PM3 and for substituted benzoic acids solvated by one water molecule, shown as two orthogonal views for (*c*) AM1 and (*d*) PM3

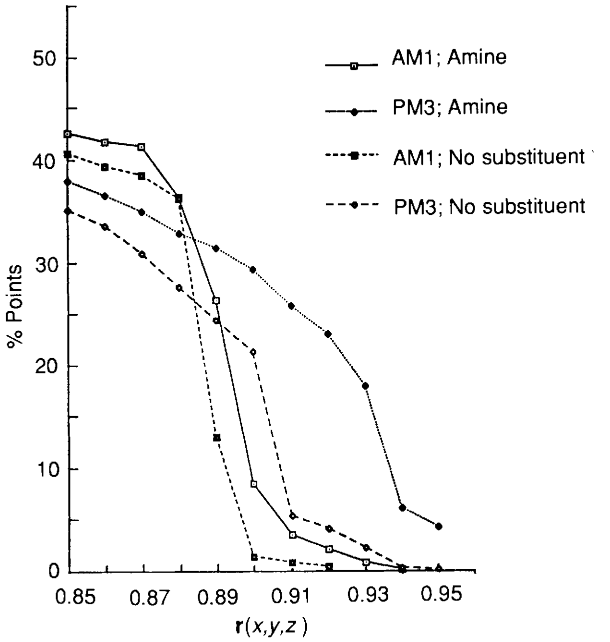

Fig. 3 Distribution of rank correlation coefficients between calculated values of the MEP and the Hammett σ_p-constants for substituted anilines, expressed as a percentage having value $\geqslant \mathbf{r}(x,y,z)$

points shows a correlation of $\mathbf{r}_s > 0.95$ at the PM3 level (Fig. 1). The spatial distribution of points reveals significant differences in the AM1 [Fig. 2(*a*)][10] and PM3 [Fig. 2(*b*)] methods at the $\mathbf{r}_s = 0.90$ level, particularly in the carboxy oxygen lone pair regions. As the value of $\mathbf{r}_s$ increases, the contours tend to focus onto the carboxy group specifically, thus providing a theoretical descriptor for the site of deprotonation. When the carboxy group is solvated with a water molecule, the PM3 geometry shows a small out of plane twisting of the CO_2 group and a significantly non-planar orientation of the water, the latter result being supported by *ab initio* calculations on related systems.[11] The effect of including this specific hydrogen bonding interaction is to significantly increase the percentage of values of $\mathbf{r}_s(x,y,z)$ in the region 0.9–0.93 for both methods (Fig. 1), although the three-dimensional distribution of these points is significantly different (Fig. 2). The addition of this solvent water enhances the maximum PM3 correlation to $\mathbf{r}_s(x,y,z) = 0.96$. That this occurs in a region of space occupied by the water molecule and the acid hydoxy lone pairs seems structurally entirely reasonable. In the region $\mathbf{r}_s(x,y,z) = 0.94$–$0.95$, the inclusion of a water molecule at the PM3 level in fact decreases the number of rank correlated points. This is because changes in the geometry of the coordinated water as a function of the substituent mean that within our coordinate system† any given fixed spatial coordinate (x,y,z) does not bear exactly the same relationship to the carbonyl function across the series of substituents. This implies the rank correlation technique, in common with parametric methods such as CoMFA, is less useful for systems where geometry or conformation is particularly sensitive to the remote substituent. Rank correlation of the MEPs derived from a monopolar expansion of the ZDO derived point charges results in <2% of the grid points with $\mathbf{r}_s(x,y,z) > 0.90$, emphasising the importance of evaluating the potential quantum mechanically.[4]

For the *p*-substituted amines, we note a significant reduction in the percentage of points with $\mathbf{r}_s(x,y,z) > 0.90$ when using AM1 (Fig. 3). With some substituents ($X = NO_2$) the nitrogen is planar, whereas with others (X = MeO) it is highly pyramidal, and the geometry changes mean that the potential at any one point (x,y,z) over the range of molecules is not in the same position with respect to the amine group. Hence, small effects at long range tend to suffer and correlation is

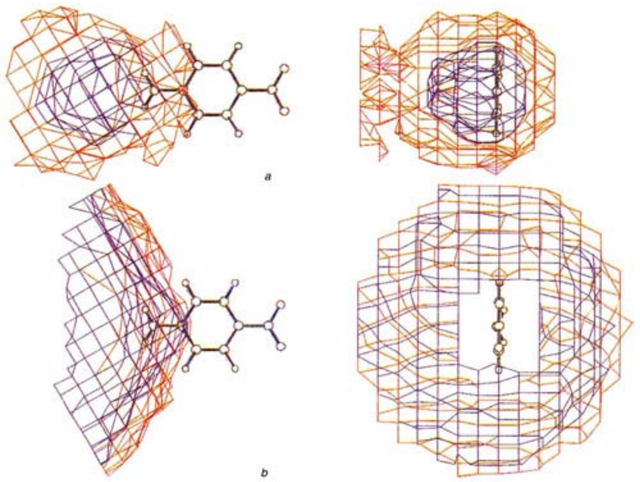

Fig. 4 Three dimensional distribution of values of **r**(*x*,*y*,*z*), contoured at **r** = 0.90 (red) and **r** = 0.92 (blue) for substituted anilines, shown as two orthogonal views for (*a*) AM1 and (*b*) PM3

reduced. If the correlation is performed using the position of the HNH group as a common geometrical descriptor,† the correlation is preserved, indicating the importance of the choice of such descriptors. With PM3, the nitrogen (incorrectly)[12] remains pyramidal over the substituent range, preserving correlation at long range. The spatial distribution of points at both the AM1 and PM3 levels is associated not only with the nitrogen lone pair region, but also with the in-plane region, illustrating that both mesomeric and inductive effects play a role (Fig. 4). If the common substituent (CO_2H or NH_2) is omitted entirely, a significantly smaller number of points show a high rank correlation (Fig. 3), confirming the important electronic role for the protonating group in concentrating the correlation with the Hammett σ_p-parameters.

We believe that rank correlation analysis between calculated electrostatic potentials and related experimental properties such as substituent or *e.g.* binding constants may provide a facile three-dimensional focus on those regions of the molecule which are most closely related to the measured property, information which for a more complex system may not be readily available. Such studies on *e.g.* binding constants will be reported in future papers.

We thank the SERC and the James Black Foundation for a Studentship (to S. M. G.) and the SERC for computing equipment grants.

Received, 9th August 1991; Com. 1/04177K

References

1 E. Scrocco and J. Tomasi, *Topics in Current Chemistry*, 1973, **42**, 95; *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, ed. P. Politzer and D. G. Truhlar, Plenum, New York, 1981.
2 P. Nagy, K. Novak and G. Szasz, *J. Mol. Struct. (Theochem.)*, 1989, **60**, 257.
3 P. K. Weiner, R. Langridge, J. M. Blaney, R. Schaefer and P. A. Kollman, *Proc. Natl. Acad. Sci. USA*, 1982, **79**, 3754; C. Petrongolo, *Gazz. Chim. Ital.*, 1978, **108**, 445.
4 F. J. Luque, F. Illas and M. Orozco, *J. Comp. Chem.*, 1990, **11**, 416; M. Orozco and F. J. Luque, *J. Comp. Chem.*, 1990, **11**, 909; G. G. Ferenczy, C. A. Reynolds and W. G. Richards, *J. Comp. Chem.*, 1990, **11**, 159.
5 B. H. Besler, K. M. Merz and P. A. Kollman, *J. Comp. Chem.*, 1990, **11**, 431.
6 H. S. Rzepa, M. L. Webb, A. M. Z. Slawin and D. J. Williams, *J. Chem. Soc., Chem. Commun.*, 1991, 765; P. N. Jagg, P. F. Kelly, H. S. Rzepa, D. J. Williams, J. D. Woollins and W. Wylie, *J. Chem. Soc., Chem. Commun.*, 1991, 942.
7 G. R. Marshall and R. D. Cramer III, *Trends Pharmacol. Sci.*, 1988, **9**, 285 and references cited therein; M. Clark, R. D. Cramer, D. M. Jones, D. E. Patterson and P. E. Simeroth, *Tetrahedron Comput. Methodol.*, 1990, **3**, 47; K. H. Kim and W. C. Martin, *J. Org. Chem.*, 1991, **56**, 2723.
8 C. Hansch, A. Leo and R. W. Taft, *Chem. Rev.*, 1991, **91**, 165.
9 C. Spearman, *Am. J. Phsych.*, 1904, **15**, 88; W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1988.
10 Displayed using SYBYL 5.41, available from Tripos Associates Inc. and with stereoscopic viewing on Evans and Sutherland PS390 and ESV10 systems.
11 H. S. Rzepa and M. Yi, *J. Chem. Soc., Perkin Trans. 2*, 1990, 943; H. S. Rzepa and M. Yi, *J. Chem. Soc., Perkin Trans. 2*, 1991, 531.
12 I. Juranic, H. S. Rzepa and M. Yi, *J. Chem. Soc., Perkin Trans. 2*, 1990, 877.