Journal of Computer-Aided Molecular Design, 5 (1991) 41–54
ESCOM

J-CAMD 103

On the representation of electrostatic fields around ab initio charge distributions

Sarah L. Price[a,*] and Nigel G.J. Richards[b]

[a]*Department of Theoretical Chemistry, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.*
[b]*Department of Chemistry, The University, Southampton SO9 5NH, U.K.*

Received 5 June 1990
Accepted 30 June 1990

Key words: Electrostatic fields; Distributed multipoles; Hydrogen bonding; Molecular graphics

SUMMARY

We compare two methods (Mulliken charges and a distributed multipole analysis, DMA) of representing an ab initio charge distribution for calculating the electrostatic field and potential outside the molecule, using pyrimidine and the RNA base uracil as examples. This is done using a 3-D graphical display of the electrostatic fields, which, when used with real-time rotation, zooming and clipping, has many advantages for qualitatively assessing the electrostatic interactions of a molecule. The errors involved in using Mulliken point charges may be of similar magnitude to the total electrostatic field in regions which are important in recognition processes. The DMA representation automatically includes the anisotropic electrostatic effects of non-spherical features in the charge distribution of each atom, and yet the displayed electrostatic fields around the atoms which have lone-pair density do not show marked anisotropy.

INTRODUCTION

It is widely believed that electrostatic forces dominate molecular recognition processes, provided steric constraints are not violated [1]. Qualitative arguments can be based upon the electrostatic properties of certain functional groups. For example, carbonyl oxygen atoms are negatively charged and therefore interact with protons to form hydrogen bonds. However, quantitative models must reflect the nature of directly bonded functional groups which modulate an atom's charge distribution through short-range inductive effects [2]. More distant functionality can also have a significant effect by polarising the charge distribution. Since both effects are naturally represented in an ab initio charge distribution, the best source of information about the electrostatic properties of a molecule remains a high-quality wavefunction.

*To whom correspondence should be addressed at: Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

Once the molecular wavefunction for a specific compound has been calculated, there remain two major problems in using the charge distribution to study the electrostatic interactions of the molecule, and thus tackling problems in molecular design. Firstly, a representation of the ab initio charge density must be chosen so that the molecular electrostatic potential (MEP) and its associated field can be evaluated rapidly at many points, without losing accuracy. Secondly, this information needs to be displayed in a form which can be easily interpreted. This is a particular problem for large unsymmetrical and stereochemically complex structures, such as the immunosuppressant cyclosporin A [3]. Similarly, a study of the binding interactions between influenza virus hemagglutinin and sialic acids [4] required a representation which indicated the degree of similarity of electrostatic fields in defined regions within different molecules. Such similarity may not be obvious from a comparison of 2-D contour plots, or by examining the electrostatic potential colour-coded onto predetermined molecular surfaces, such as the van der Waals surface [5] or one defined by a fixed ratio of the van der Waals radii [6], because the chosen surface may not pass through the critical region.

In most molecular modelling studies, especially those involving molecular dynamics [7], the charge distribution is represented by atomic point charges located at the nuclear positions. In many modelling packages these partial charges are obtained by a Mulliken analysis of an ab initio wavefunction [8]. Other methods for obtaining partial charges are becoming popular, including fitting the charges directly to the ab initio MEP, at defined grid points [9]. However, any point-charge model which only uses atomic sites, represents the molecule as a superposition of spherical charge densities. This implies that, upon bonding, valence-electron density is removed from electropositive and relocated on electronegative atoms so that all of the atoms possess spherically symmetric charge densities. This is contrary to almost all qualitative theories of chemical bonding [10] which predict that the valence-electron distribution about a given atom is generally far from spherical and will often include features such as lone pairs and π-electron density. A more complete model of the charge distribution [11–15] would be a set of point multipoles (charge, dipole, quadrupole, etc.), at every atom. The higher multipole moments enable the model to describe the effects of non-spherical features in the charge distribution on intermolecular electrostatic interactions. Hence any mechanism responsible for the alignment of two, or more, molecules with respect to each other, such as hydrogen bonding [16], and which is influenced by electrostatic forces arising from such features as lone-pair electron density, is best modelled using distributed multipole models.

Computer graphics provides chemists with a tool for the rapid analysis of large, multifunctional molecular structures, such as proteins [17]. Standard graphical representations of molecules only duplicate physical models (Dreiding or CPK). Although this has allowed the chemist to manipulate structures in three dimensions, which is crucial to the analysis of spatial relationships, it has also tended to reinforce traditional ideas about molecular structure. Given recent developments in workstation technology, it is now possible to use real-time graphics to study other molecular properties, such as the electrostatic potential or hydrophobicity.

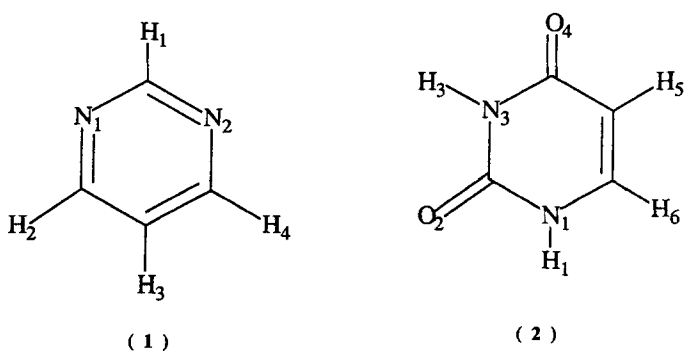
In the area of electrostatics, most work has concentrated on colour coding, or contouring, the potential onto predefined molecular surfaces or planes [6, 7, 18]. These representations are informative, and readily encapsulated in a photograph. However, they do not exploit the ability of a molecular graphics system to represent a property in all space, which can then be examined interactively using the rotation, zooming and clipping facilities. A 3-D display has many advantages

for a preliminary assessment of the electrostatic properties of a molecule, and can be used to determine which regions require detailed examination by more conventional methods. It is also essential to represent a property so that the links between it and the functional groups are apparent. This is often not the case when the electrostatic potential is colour-coded onto a surface, unless the molecular shape is so distinctive that the type of the atom enclosed by each sphere is obvious from all viewpoints.

The electrostatic field is a natural representation of the electrostatic properties of a molecule, and can be displayed graphically as a series of vectors located at defined points in space. In this way, qualitative rationalisations for the interaction of DNA with ethidium [19], and the surprisingly high rate of reaction of superoxide dismutase have been obtained [20]. In this work we develop the use of the vector representation of the electrostatic field about a molecule, visualised as a Dreiding model together with its van der Waals surface. Since the field at a given grid point can be manipulated using vector algebra, this representation is particularly useful, not only for examining this property in three dimensions, but also for investigating the differences between fields calculated using different models of the ab initio charge distribution. We illustrate this point by comparing the fields calculated from a distributed multipole model and the corresponding Mulliken charges for pyrimidine and uracil. The key difference between these analyses of the ab initio charge density is the representation of non-spherical features in the atomic charge distributions. Thus, the effects of lone pairs, and π-electrons, on the electrostatic field around a molecule are highlighted by our graphical approach.

MATERIALS AND METHODS

The planar compounds, pyrimidine and uracil (Fig. 1) were selected as test molecules for this study, in order to assess effectiveness of the arrow representation of their electrostatic properties. This choice also removed the need to consider the effects of conformation upon the charge distribution, with all of the attendant complications. If the representation was not useful for these symmetric structures, then the extension of this approach to stereochemically complex macromolecules would be worthless. In addition, this choice meant that relatively few different views of the molecule conveyed most of the information which could be obtained at the molecular graphics



(1) (2)

Fig. 1. Numbering system for pyrimidine (**1**) and uracil (**2**).

terminal. The coordinates for pyrimidine were based on crystal structure data [21], after adjustment to give the molecule exact C_{2v} symmetry [22]. The uracil geometry was based on average crystal structure dimensions [23], with the addition of hydrogen atom coordinates calculated using standard bond lengths and angles. All SCF calculations were performed with a double-zeta quality basis set [24] using the CADPAC suite of programs [25]. The Mulliken and the distributed multipole analysis (DMA) of the wavefunction were obtained using the command options PROPERTIES and MULTIPOLES. Both the latter analyses were performed at an insignificant computational cost relative to the calculation of the wavefunction. Both analyses of the molecular charge density $\rho(\mathbf{r})$ use the charge density matrix ρ_{ij} which can be defined in terms of the Gaussian primitive functions φ_i which make up the basis set used to calculate the wavefunction:

$$\rho(\mathbf{r}) = \sum \rho_{ij}\, \varphi_i(\mathbf{r})\varphi_j(\mathbf{r}) \tag{1}$$

In Mulliken analysis [8], the contribution to the charge density when both φ_i and φ_j are centered upon a given atom is represented by a charge at that atom, and when the orbitals are based upon different atoms ($i \neq j$), the charge is split equally between atoms i and j. In contrast, the DMA [11] recognises that a term in Eq. 1 can only be represented exactly by just a charge if both φ_i and φ_j are s orbital Gaussians. The overlap $\varphi_i\, \varphi_j$ from an s- and p-orbital, is equivalent (outside the charge density) to a point charge and dipole, and that from two p-orbitals to a charge, dipole and quadrupole, located at the overlap centre. If both orbitals are centered on the same atom, the overlap centre is at that atom. Otherwise it is at a point determined by the exponents and centres of the two Gaussians. The number of overlap centres arising from any reasonable basis set is so large, that evaluating the electrostatic potential from these sets of multipoles is almost as prohibitively expensive as integrating over the charge distribution. Hence it is necessary to choose a much smaller set of interaction sites. In this study, sites were chosen to be at every nuclear position. Clearly, extra sites can be added for higher accuracy, just as sites may be removed if simpler models are adequate for the modelling problem. Any overlap contribution which is not already at one of the chosen sites, was represented at the nearest site by an infinite series of multipoles. Since the formula determining the shifting algorithm depended on the distance d between the overlap centre and the interaction site, the multipole series on each site converged rapidly after the quadrupole moment, as d was small, and the molecules were described by sp basis sets. The electrostatic potential, field, or interaction energy, was readily computed from the DMAs of both molecules, using the analytical expressions for the electrostatic energies and forces between point multipoles [26], as programmed in ORIENT [27]. In this work, each component of the electric field was calculated from the interaction energy of the molecule with a point dipole, using all terms in the series expansion up to R^{-5}. Thus the multipole moments up to octupole (Q_{3k} or Ω) contributed to the electrostatic field at each point. The points used were a cubic grid of 0.5 Å spacing, excluding any points which either lay within, or more than 3 Å distant from, the van der Waals surface of the molecule. This molecular surface was defined using standard Pauling radii of 1.5 Å for N, 1.4 Å for O and 2.0 Å for 'united' carbon atoms [10]. Since the hydrogen atom had no explicit van der Waals radius, the grid could sample space close to the hydrogen atoms in NH bonds, as occurs in hydrogen bonding.

The results of the calculations were displayed, using a modified version of the MacroModel [28] program, V2.5, upon an Evans & Sutherland PS390 hosted by a MicroVAX II computer. Arrows,

TABLE 1
COLOUR CODE FOR THE MAGNITUDE OF THE VECTORS

Colour	Lower limit in V/Å	in kT/Debye[a]
Purple	0.247	2
Blue	0.493	4
Green	0.740	6
Yellow	0.987	8
Orange[b]	1.234	10
Red[b]	1.480	12

[a]Interaction energy of a dipole of 1 Debye at 298 K.
[b]Unfortunately red and orange are not well resolved in the graphics.

representing the electrostatic field, were placed such that their tails were located at the defined grid points, and were coloured according to the field magnitude. In contrast to previous work, the arrowlength was fixed at a size equal to the grid-point spacing, and therefore had no physical significance. However, this introduced a 3-D perspective when the vectors were displayed on the graphics terminal, which was enhanced by the apparent variations in length as the structure was rotated. The direction of a given arrow corresponded to that in which a positive charge would move under the influence of the field, as in the normal definition of an electrostatic field ($\mathbf{E} = -\nabla\varphi(\mathbf{r})$). Arrow colour was determined by the field strength as computed from the energy of interaction between the field and a molecule possessing a dipole moment of 1 Debye. Limits for the range of energies represented by each colour were set in multiples of k_BT, where k_B is the Boltzmann constant, and the temperature T was chosen to be 298 K (Table 1). Since we were interested in the role of electrostatics in orienting molecular species during recognition events, only vectors representing field interaction energies greater than 2 k_BT/Debye (0.247V/Å) were displayed. This significantly reduced the number of displayed vectors from the several thousand computed around each structure, and thus picked out the regions of space where electrostatic effects would be important. For graphics processors, such as the PS390, which are usually limited by available local memory, the data reduction also facilitates the display of electrostatic information about complex molecules, such as proteins or DNA. In addition to giving the coordinates and electrostatic models, a simple FORTRAN-77 subroutine for obtaining the vector information required for rendering arrows will be deposited with this journal in machine-readable form, and is available on request.

RESULTS

Pyrimidine

The systematic nature of the distributed multipoles of the azabenzenes has already been described [22], using a DMA with sites located only upon the carbon and nitrogen atoms. In the all-atom DMA used here, the previous qualitative picture of the charge distribution is unaffected since all multipole moments on the hydrogens are small. Short-range inductive effects are seen as both nitrogens possess a partial charge of ca. −0.6e, having taken electron density from both adjacent carbons. This results in a high positive charge of ca. 0.5e on C1, in contrast to the small

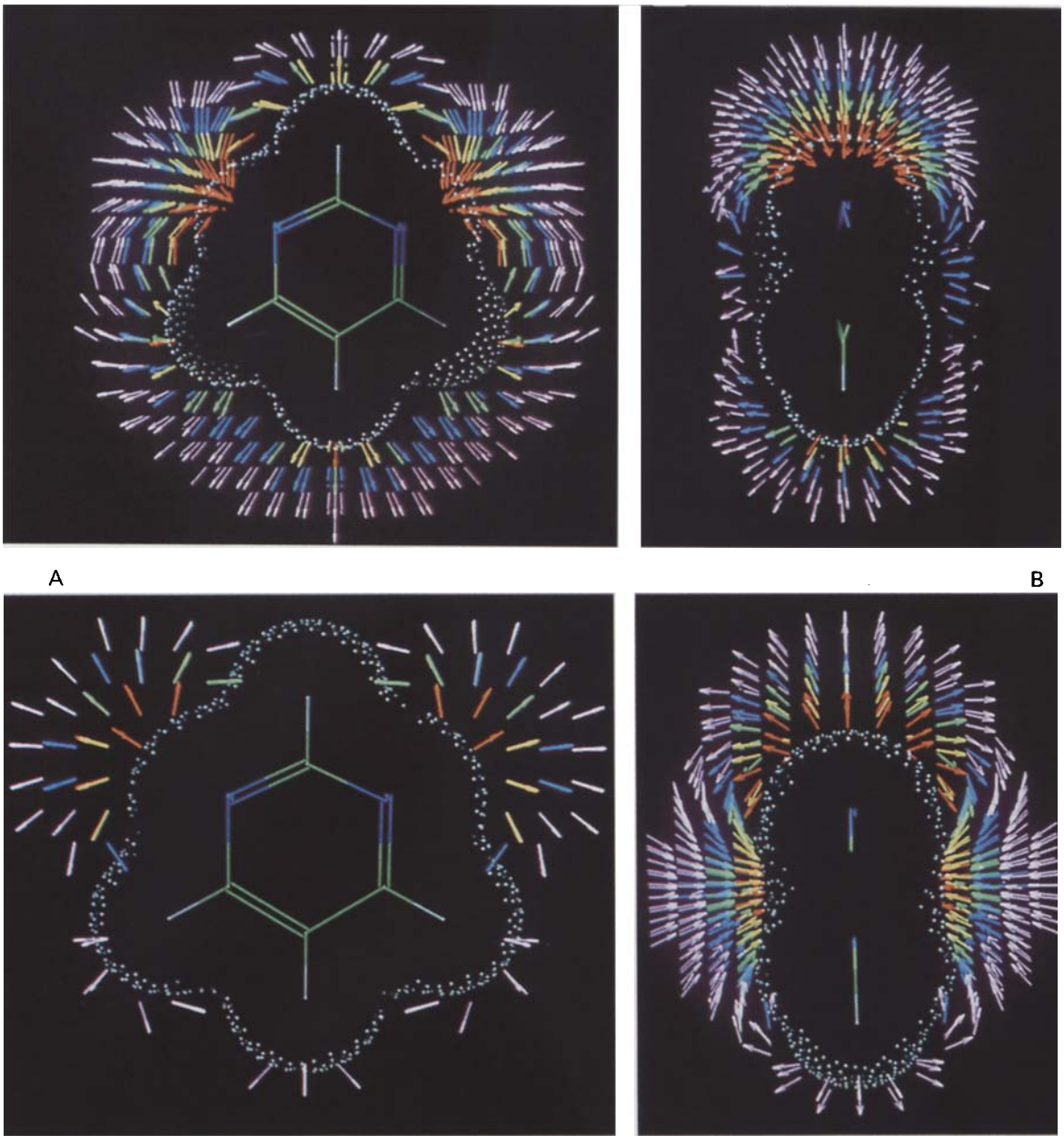


Fig. 2. Electrostatic field representations for pyrimidine. (A) $\mathbf{E}_{DMA}$: Electrostatic field calculated from the DMA. The view is almost perpendicular to the ring, with a slight twist to enhance the sense of depth and clipping. (B) $\mathbf{E}_{DMA}$: View from the side of the ring, clipped to the plane through the centre of the molecule. (C) $\mathbf{E}_{Diff}=\mathbf{E}_{Mull}-\mathbf{E}_{DMA}$: Errors in the field when calculated from the Mulliken charges rather than DMA of the same wavefunction. This view is perpendicular to the molecule, and severely clipped to the plane of the molecule. (D) $\mathbf{E}_{Diff}=\mathbf{E}_{Mull}-\mathbf{E}_{DMA}$: Viewed from the side of the molecule, clipped to the plane through the centre of the ring. The white dot surface represents the van der Waals envelope, using the default radii of the MACROMODEL program version 2.5 [28]. The atoms in the Dreiding representation of the molecule are colour-coded red for O, blue for N, green for C and white for H.

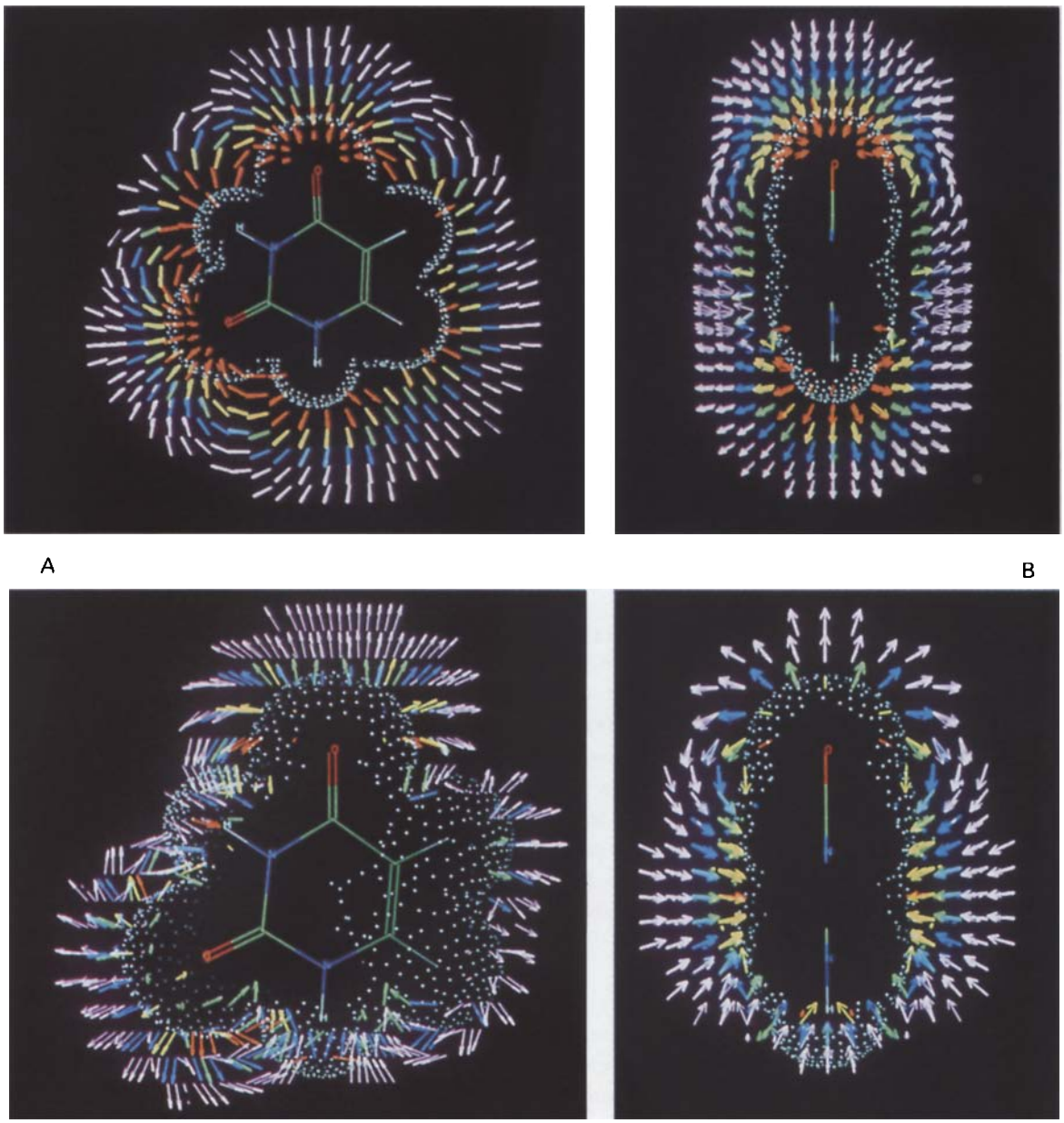


Fig. 3. Electrostatic field representations for uracil. (A) $\mathbf{E}_{DMA}$: Electrostatic field calculated from the DMA. The view is almost perpendicular to the ring, with a slight twist to enhance the sense of depth, and clipping. (B) $\mathbf{E}_{DMA}$: View from the side of the ring, clipped to the plane through the centre of the molecule. (C) $\mathbf{E}_{Diff} = \mathbf{E}_{Mull} - \mathbf{E}_{DMA}$: Errors in the field when calculated from the Mulliken charges rather than DMA of the same wavefunction. This view is approximately perpendicular to the molecule, with a slight twist and slight clipping. (D) $\mathbf{E}_{Diff} = \mathbf{E}_{Mull} - \mathbf{E}_{DMA}$: Viewed from the side of the molecule, clipped to the plane through the centre of the ring. The white dot surface represents the van der Waals envelope, using the default radii of the MACROMODEL program version 2.5 [28]. The atoms in the Dreiding representation of the molecule are colour-coded red for O, blue for N, green for C and white for H.

charge of −0.01e on C3. The atomic dipole moments reflect the nitrogen lone-pair density and refine the picture of charge transfer along the bonds, and π-electron density above, and below, the ring is reflected in the atomic quadrupole moment component Q_{20}. The DMA of pyrimidine therefore shows the anisotropies in the molecular charge distribution which would be expected from chemical bonding theory.

The electrostatic field about pyrimidine, calculated from the DMA, is shown using the arrow representation in Fig. 2. The effectiveness of colour in directing the eye to the regions where the electrostatic field around the molecule is strongest, and the 'flow' associated with the field lines, are both striking. Addition of a standard white dot representation of the van der Waals surface emphasises the range over which the electrostatic field might have a strong orientating effect at room temperature, and provides a sense of the degree of clipping used in the picture.

The largest potential gradients are close to the nitrogen atoms, but, significantly, the field is not highly directional in these regions. Thus the lone-pair density does not appear to be acting as a pseudoatom, or to have the degree of directionality predicted by naive concepts of lone pairs in the σ-framework. The field gradients about the CH groups are quite large and not too dissimilar. Indeed, the largest gradients are found in the region of C3. This is noteworthy, since the carbon atoms bear very different charges; C1 bearing a partial charge of similar magnitude to those observed for both nitrogen atoms. This emphasises that the local field about any atom includes major contributions from other atoms in the molecule because of the long range of the electrostatic forces.

The difference field ($\mathbf{E}_{\text{Diff}} = \mathbf{E}_{\text{Mull}} - \mathbf{E}_{\text{DMA}}$), computed by simple vector subtraction of the fields due to the Mulliken charges and the DMA respectively, is also shown in Fig. 2. Identical field strengths were used in defining the colouring, so that the magnitude of the errors between the electrostatic field calculated from Mulliken charges is clearly of the same order as the total field. The large magnitude difference vectors show high anisotropy in regions about the nitrogen atoms. Significant errors are also observed above the aromatic ring, although the difference field about the CH groups in the plane of the molecule is small, with only purple vectors very close to the van der Waals surface. This correlation of the largest errors with the non-spherical features of the nitrogen lone-pair density and the π-electron ring reinforces the notion that the major deficiency of the Mulliken partial charge analysis is its inability to represent non-spherical features in the charge density.

Uracil

Strong, short-range, inductive effects are also observed in the DMA of uracil. Again, the multipoles associated with each atom are dependent on the nature of functional groups which are directly bonded. Thus the two oxygen atoms have fairly similar charges (−0.849e for O2 and −0.830 for O4), whereas those upon their adjacent carbon atoms differ markedly (1.162e for C2 vs. 0.921e for C4). Presumably this is due to the effects of the two nitrogen atoms flanking C2 in the ring. Incorporation of higher multipole moments significantly refines the spherical picture of the charge distribution around each atom given by a charge-only model. All the atoms in the six-membered ring have large in-plane octupole moment components Q_{33s}, representing the directionality of the charge density along the three bonds. The other in-plane multipole moments then reflect differences between the three bonds.

The electrostatic field vectors calculated from the DMA charge distribution (Figs. 3A and B)

indicate that the highest magnitude fields are associated with the NH and carbonyl groups, and, again there is a clear impression of the direction of charge motion about the molecule. The magnitude of the field vectors close to the carbonyl oxygens does not depend strongly upon direction. Significant fields are observed towards the centre of the ring, and close to the protons on the C=C double bond even though the partial charges determined for these protons are less than 0.07e. The field above the molecular plane about the centre of the ring is stronger for uracil than for pyrimidine (contrast Figs. 2B and 3B).

The field difference vectors ($\mathbf{E}_{Diff} = \mathbf{E}_{Mull} - \mathbf{E}_{DMA}$) are often of similar magnitude to the original fields, exceeding interaction energies of 12 k_BT/Debye (Figs. 3C and D). However, fewer vectors are of sufficient magnitude to be displayed; there are only 1299 $\mathbf{E}_{Diff}$ vectors, in contrast to 3405 field vectors $\mathbf{E}_{DMA}$, which are greater than 2 k_BT/Debye. While large errors in the electrostatic field around the 'sides' of the carbonyl oxygens are observed, the errors at the end of the CO group are small. There are also large differences in the fields above the centre of the ring. Although the Mulliken charges seem to be able to represent the field around the C=C double bond in the plane of the molecule, there are significant errors above the bond. In general, the region of substantial field differences due to the use of Mulliken charges is smaller for uracil than for pyrimidine. In both cases, however, this volume extends for a significant distance above the ring (Figs. 2D and 3D).

DISCUSSION

Representation of the charge distribution

This graphical representation undoubtedly shows the marked differences between the electrostatic fields calculated from Mulliken charges and from a more complete analysis (DMA) of the same ab initio charge distribution. The observed anisotropy of the difference vectors accords with the inability of Mulliken charges to represent non-spherical features in the charge density. To confirm that the errors primarily arise from the use of just point charges, rather than the method of determining the charge, differences between the electric fields computed from just the charge component of the DMA and the full DMA, respectively, were displayed. These sets of difference vectors, which arise directly from the contribution of anisotropic multipole moments, were qualitatively similar to the $\mathbf{E}_{Diff}$ ($\mathbf{E}_{Mull} - \mathbf{E}_{DMA}$) maps. In the case of pyrimidine, the omission of the anisotropic multipole moments in the DMA resulted in the display of 1168 difference vectors of energy greater than 2 k_BT/Debye with 181 (15%) having an electrostatic field magnitude corresponding to 6 k_BT/Debye at T=298 K. The corresponding figures for uracil were 1299 and 255 (20%), respectively. For pyrimidine, the largest errors arising from the omission of anisotropic multipole moments were around the nitrogen atoms, with slightly more red arrows in the direction of C1 than C2 or C4. There was also a red arrow along the C2H and C4H bonds, and some green arrows above the aromatic ring. Everywhere else no field differences corresponding to energies much higher than thermal motion were present (blue and purple). For uracil, the largest differences due to the higher multipoles were associated with the C=C double bond and its attendant protons. Resultant fields $\mathbf{E}_{Diff}$ (yellow and green vectors) of smaller magnitude were located on either side of the carbonyl oxygen (O4). Inevitably, the errors in any charge-only electrostatic model vary with the method used to assign the charges. However, these results show graphically that an atomic point-charge model is intrinsically limited near the van der Waals surface. This conclusion is consistent with recent work [9, 29, 30] on deriving point-charge models by fitting directly to mo-

lecular electrostatic potential, evaluated by integration over the wavefunction at a large number of grid points located outside the van der Waals surface of the molecule. Even though this approach provides the best available model consistent with the assumption of point charges located at the nuclear position (spherical atoms), the residual errors in the fitting can be quite significant. More elaborate electrostatic models, which include atomic, or bond, dipole moments, and atomic quadrupoles have also been derived by fitting to a potential grid [31]. However, an important limitation of methods of deriving electrostatic models by fitting, which is not shared by a wavefunction analysis such as the DMA, is that the computational cost of calculating the potential at sufficient grid points by integration becomes prohibitively expensive for large basis sets and molecules containing many atoms.

An important problem in the use of ab initio based electrostatic models is the choice of basis set. Unfortunately, those in current use for organic molecules are far from the Hartree–Fock limit. In addition, calculations on small molecules suggest that electron correlation will also affect the electrostatic properties derived from an ab initio wavefunction significantly [32]. We therefore examined the differences in the electrostatic field about uracil calculated from two different DMAs, obtained from the double-zeta basis wavefunction and a 3-21G basis wavefunction [33]. The displayed difference field ($\mathbf{E}_{\text{Basis}} = \mathbf{E}_{\text{3-21G}} - \mathbf{E}_{\text{DZ}}$), contained far fewer vectors than $\mathbf{E}_{\text{Diff}} (= \mathbf{E}_{\text{Mull}} - \mathbf{E}_{\text{DMA}})$. Only 291 $\mathbf{E}_{\text{Basis}}$ field vectors were above the cutoff strength of 2 k_{B}T/Debye in contrast to 1299 for $\mathbf{E}_{\text{Diff}}$, and all were located at grid points close to the molecular surface. Furthermore, only 20 $\mathbf{E}_{\text{Basis}}$ vectors had a magnitude of greater than 6 k_{B}T/Debye (7%, compare 20% for $\mathbf{E}_{\text{Diff}}$), and these were at the van der Waals surface of the amide protons and carbonyl oxygens. Nevertheless, there are significant differences between the charge densities calculated using the two basis sets, which clearly illustrates the problem in selection of a given basis set. The double-zeta basis calculation predicted a total dipole moment of 5.615 D for uracil, significantly larger than the 5.061 D computed using the 3-21G basis set. An experimental value of 4.16 ± 0.04 D has been reported for the dipole moment of uracil in dioxane [34], but considerable uncertainty exists as to the influence of solvation effects. These values can be contrasted with other literature values of 5.42 D obtained using a 4-31G basis set [35] and 4.8 D for an optimised geometry calculated with the 3-21G basis set [36]. Whatever the exact relationship of the calculated dipole moment to reality, the graphical representation clearly showed that the errors in using Mulliken charges rather than a DMA were larger than those due to differences between basis sets, although literature debate upon the latter problem is far more extensive. The choice of model representation for an ab initio charge distribution is, therefore, at least as important as the choice of basis set, and these results add to the already considerable evidence [29, 30, 37] that Mulliken charges are unsuitable for representing a molecular charge distribution. Whether a better point-charge model, such as potential-derived charges, will be any better should be related to the nature of the simulation because the importance of anisotropic terms is dependent upon three main factors. These are the actual distribution of the valence electrons, whether a large portion of the van der Waals surface of the most anisotropic atoms is exposed, and how close to the molecule the electrostatic potential is sampled. In most cases, a more complete representation, such as a DMA, will probably be required to ensure that the electrostatic model represents the ab initio charge density well.

Implications for molecular recognition

Given that the DMA model provides an accurate picture of the electrostatic field which in-

cludes contributions from non-spherical features such as lone pairs, it is noteworthy that the field localised about the nitrogen atoms in pyrimidine, and the carbonyl oxygens in uracil, did not exhibit any marked anisotropy. However, the lack of any 'rabbit ears' of electron density round carbonyl oxygens, or a lone pair which resembles a 'baseball bat' protruding from an sp^2 nitrogen atom (as lone-pair density is often depicted) is quite reasonable since such electron density appears as a squat lobe when observed experimentally for pyrazine in density difference maps [38]. Furthermore, the fact that the electrostatic field about oxygen and nitrogen atoms can be reasonably isotropic (within the scale of this representation) is not inconsistent with the DMAs of these atoms possessing significant anisotropic multipole moments, because electrostatic interactions operate over such long range. Hence, there is almost certainly significant cancellation between the electric fields arising from the lone-pair density, and contributions from the multipoles on neighbouring atoms.

The lack of marked anisotropy corresponding to lone-pair density, as described above, has significant implications for the simulation of hydrogen bonding between the functionality in pyrimidine and uracil and suitable donors. A statistical analysis [39] of the geometries of C=O···H−N non-covalent interactions, in small molecule crystal structures, has indicated a statistical preference for hydrogen bonds to form with directional dependencies corresponding to those of conventionally viewed sp^2 lone pairs. This result could be rationalised as resulting from a strong directionality of the electrostatic field about carbonyl oxygens, such as those in uracil, but such anisotropy has not been observed using this field representation (Fig. 3A). Surprisingly, this is not necessarily in conflict with the results of the crystal data analysis. Nor does it suggest hydrogen bonding directionality is not electrostatic in origin, since the sterically accessible minima in the electrostatic energies of a few dozen complexes containing C=O···H−N bonds [40], calculated using DMA models, showed the same structural preferences as observed in the crystal structures. However, examination of the electrostatically favoured structures of the complexes [40] revealed that a planar, linear hydrogen bond in the lone-pair direction usually corresponded to a second favourable contact between the molecules.

Why visualise and use electrostatic fields?

Recent advances (particularly over the past 18 months) in the price/performance ratio of graphics workstations, have made it possible to render and manipulate solid objects in three-dimensional space in real time. Given that colour coding the MEP onto a number of predefined surfaces (which might be visualised using clipping methods, or transparency) remains a popular approach to studying molecular electrostatics, how useful is the arrow representation described here? This question is especially pertinent to planar molecules, such as pyrimidine and uracil, given the simple nature of their van der Waals surface, and their molecular symmetry. A noticeable benefit is the introduction of a dynamic element into the visual interpretation of molecular electrostatics, as the association of directionality with physical movement is an intuitive notion familiar from high-school physics. We have found that the eye naturally follows the arrows from regions of high negative to high positive potential, immediately providing insights into the motion of a molecular dipole about the structure, and thereby generating models for docking pathways which might be followed during intermolecular recognition. In addition, since the grid spacing is known, examination of the arrow colouring enables a rapid assessment of the distance from the van der Waals surface of the molecule, through which the electrostatic potential might play a significant role in

mediating intermolecular association. Such properties of the representation have been essential to modelling the role of electrostatics in controlling the orientation of substrate within the binding pocket of superoxide dismutase [20] and in the recognition of oligosaccharides containing *N*-acetylneuraminic acid by influenza virus haemagglutinin [5].

The overwhelming benefit of producing a simple representation of the electrostatic field, however, is the ability to manipulate it interactively within 3-D space upon the graphics processor. Even for these planar structures one learns less from a few 2-D snapshots than from being able to rotate, clip and zoom the 3-D representation interactively. The ability to search for key regions around the molecule using interactive graphics assumes greater importance for larger molecules, which have little symmetry. The interactive clipping is essential because such a large number of vectors have to be generated to describe the 3-D electrostatic field about the molecule, that it is necessary to be able to focus on different regions. On a more mundane level, colour coding makes it easy to spot the regions of highest interest. These can be analysed in more detail using contour plots, or surfaces if required.

The simple mathematical nature of the electrostatic field at a series of grid points about complex structures has additional computational advantages in the areas of QSAR and molecular design. As part of work aimed at probing the essential molecular features of the interaction of peptides with monoclonal antibodies [41] we have been involved in designing β-turn peptidomimetics [42]. Essential to the design of such compounds is an assessment of the similarity of the electrostatics of the analogue, which on steric grounds may play the same role in a molecular recognition process, to the original amino acid segment. In our usual approach, the dipeptide segment, in a particular β-turn conformation [43], is aligned with the proposed peptidomimetic such that isosteric functionality [44] is in a similar spatial orientation. We note that this is a non-trivial problem for structures of compatible but dissimilar shape [45]. Computation of the electrostatic field at a common set of grid points about both structures, followed by evaluation and display of the difference field is then straightforward. If the latter is weak throughout 3-D space, this is usually assumed to indicate that the peptidomimetic can perform the same biological function as the peptide segment in the appropriate β-turn conformation. Further work is also in progress to use molecular electrostatic fields to align structures in QSAR studies. In this approach, two or more, structures are initially superimposed in arbitrary orientations, and a grid constructed about them such that no grid point lies within any molecular van der Waals volume. A similarity function based upon the dot product of the various electrostatic fields at each grid point is then computed, and random translation/rotation of the structures (as rigid bodies) carried out. The function is then recomputed, and the move accepted, or rejected, using standard Metropolis criteria [46] until the optimum match of the electrostatic fields of the molecules is found.

CONCLUSIONS

The use of this simple graphical representation of the molecular electrostatic field, which can be rapidly manipulated on any current graphics processor, has unique advantages for surveying electrostatic forces, particularly those involved in molecular recognition. The use of colour automatically highlights regions of space where the electrostatic field will have a marked influence on the motion of any dipolar species, such as a water molecule, and pinpoints where this effect will be strongest. The flow of the field vectors around the molecule corresponds to the orientation of a

point dipole, and thus, when the field is strong, gives an approximate picture of the docking of polar X–H bonds into hydrogen bonding positions. Since the representation can be superimposed on a Dreiding model for the molecule, together with a dot representation of the van der Waals surface, the correlation of the electrostatic field with specific functional groups is readily established.

The extension of vector representations from the early studies of electrostatic fields in protein modelling to visualising differences in properties, has been illustrated by a detailed comparison of various models for computing the electrostatic field around a molecule. It seems clear that the inherent advantages of obtaining an electrostatic model using ab initio techniques can be lost by using a poor representation of the resulting charge density, such as Mulliken charges. The use of a DMA to give a more complete representation of this property shows that non-spherical features such as lone pairs and π-electrons *do* make a significant contribution to the electrostatic field, at least close to the van der Waals surface. Thus the use of a distributed multipole model is likely to be necessary for the realistic simulation of molecular recognition processes which sample space close to the molecular surface, such as hydrogen bonding. In turn, the implication is that the next generation of force fields for molecular mechanics simulations ought to include anisotropic terms, abandoning the assumption that molecules interact as if they were merely a superposition of spherical charge densities. This may also enable the accurate representation of hydrogen bonding interactions within the potential energy function without the use of ad hoc energy terms. Moreover, this graphical representation has emphasised that lone-pair density does not produce markedly anisotropic features like 'rabbit ears'. Electrostatic interactions are long range, and therefore the field about a given atom results from a subtle balance of anisotropic contributions from many different atoms. Thus, the qualitative picture of the relative ability of atoms to attract, or repel, other charged species, derived from chemical intuition will generally need refining using accurate quantitative models and molecular graphics.

ACKNOWLEDGEMENTS

We thank Mr. J.B.O. Mitchell for help with the ab initio calculation of uracil. S.L.P. thanks the Royal Society for a 1983 University Research Fellowship, while N.G.J.R. acknowledges the SERC (U.K.) for provision of molecular graphics equipment, and the Nuffield Foundation for additional financial support.

REFERENCES

1 Nakamura, H., Komatsu, K., Nakagawa, S. and Umeyama, H., J. Mol. Graphics., 3 (1985) 2.
2 Faerman, C.H. and Price, S.L., J. Am. Chem. Soc., 112 (1990) 4915.
3 Wenger, R.M., Angew. Chem., Int. Ed. Engl., 24 (1985) 77.
4 Richards, N.G.J., Int. J. Quant. Chem., Quant. Biol. Symp., 15 (1988) 85.
5 Thomson, C., Higgins, D. and Edge, C., J. Mol. Graphics, 6 (1988) 171.
6 Burridge, J.M., Quaredon, P., Reynolds, C.A. and Goodford, P., J. Mol. Graphics, 5 (1987) 165.
7 McCammon, J.A. and Harvey, S.C., Dynamics of Proteins and Nucleic Acids, Cambridge University Press, Cambridge, 1987.
8 Mulliken, R.S., J. Chem. Phys., 23 (1955) 1833.
9 Singh, U.C. and Kollman, P.A., J. Comput. Chem., 5 (1984) 129.
10 Pauling, L., The Nature of the Chemical Bond, 3rd edn., Ithaca Press, Cornell, NY, 1960.

11 Stone, A.J. and Alderton, M.A., Mol. Phys., 56 (1985) 1047.
12 Pullman, A. and Perahia, D., Theor. Chim. Acta, 48 (1978) 29.
13 Sokalski, W.A. and Sawaryn, A., J. Chem. Phys., 87 (1987) 526.
14 Vigné-Maeder, F. and Claverie, P., J. Chem. Phys., 88 (1988) 4934.
15 Rico, J.F., Alvarez-Collado, J.R. and Paniagua, M., Mol. Phys., 56 (1985) 1145.
16 Buckingham, A.D. and Fowler, P.W., Can. J. Chem., 63 (1985) 2018.
17 Feldman, R.J., Bing, D.H., Furie, B.C. and Furie, B., Proc. Natl. Acad. Sci. U.S.A., 75 (1978) 5409.
18 Furet, P., Sele, A. and Cohen, N.C., J. Mol. Graphics, 6 (1988) 182.
19 Dean, P.M., Molecular Foundations of Drug–Receptor Interaction, Cambridge, 1987, p. 264.
20 Getzoff, E.D., Tainer, J.A., Weiner, P.K., Kollman, P.A., Richardson, J.S. and Richardson, D.C., Nature, 326 (1983) 287.
21 Wheatly, P.J., Acta Crystallogr., 13 (1960) 80.
22 Price, S.L. and Stone, A.J., Chem. Phys. Lett., 98 (1983) 419.
23 Voet, D. and Rich, A., Prog. Nucleic Acid Res. Mol. Biol., 10 (1970) 183.
24 Dunning, T.H., J. Chem. Phys., 53 (1970) 2823.
25 Amos, R.D. and Rice, J.E., CADPAC: The Cambridge Analytical Derivatives Package, Version 4.0, Cambridge, 1987.
26 Price, S.L., Stone, A.J., and Alderton, M., Mol. Phys., 52 (1984) 987.
27 Price, S.L. and Stone, A.J., J. Chem. Phys., 86 (1987) 2859.
28 Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R.M.J., Lipton, M.A., Chang, G., Hendrickson, T., Hasel, W.F. and Still, W.C., J. Comput. Chem., 11 (1990) 440.
29 Momany, F.A., J. Phys. Chem., 82 (1978) 592.
30 Cox, S.R. and Williams, D.E., J. Comput. Chem., 2 (1981) 304.
31 Williams, D.E., J. Comput. Chem., 9 (1988) 745.
32 Amos, R.D., Chem. Phys. Lett., 113 (1985) 19.
33 Gordon, M.S., Binkley, J.S., Pople, J.A., Pietro, W.J. and Hehre, W.J., J. Am. Chem. Soc., 104 (1982) 2797.
34 Kulakawska, I., Geller, M., Lesyng, B. and Wierzchowski, K.L., Biochim. Biophys. Acta, 361 (1975) 119.
35 Eisenstein, M., Int. J. Quant. Chem., 33 (1988) 127.
36 Zielinski, T.J., J. Comput. Chem., 4 (1983) 345.
37 Price, S.L., Harrison, R.J. and Guest, M.F., J. Comput. Chem., 10 (1989) 552.
38 Feil, D. and Moss, G., Acta Crystallogr., A39 (1983) 14.
39 Taylor, R., Kennard, O. and Versichel, W., J. Am. Chem. Soc., 105 (1983) 5761.
40 Mitchell, J.B.O. and Price, S.L., Chem. Phys. Lett., 154 (1989) 267.
41 Richards, N.G.J., Hinds, M.G., Brennand, D.M., Glennie, M.J., Welsh, J.H. and Robinson, J.A., Biochem. Pharmacol., 39 (1990) in press.
42 Morgan, B.A. and Gainor, J.A., Annu. Rep. Med. Chem., 24 (1989) 243.
43 Hinds, M.G., Richards, N.G.J. and Robinson, J.A., JCS Chem. Commun., (1988) 1447.
44 Bardos, T.J., In Boschke, F. (Ed.) Medicinal Chemistry (Topics in Current Chemistry, Vol. 52), Springer-Verlag, New York, 1974, pp. 63–98.
45 Carbo, R., Leyda, L. and Arnau, M., Int. J. Quant. Chem., 17 (1980) 1185.
46 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E., J. Chem. Phys., 21 (1953) 1087.