alternation appear to be hyperconjugation and π-delocalization within the fused ring systems, although strain-induced changes in the hybridization are important in determining the geometry of highly strained **15**.

Acknowledgment. This work was supported by the National Science Foundation (Grant No. 87-21134) and the Director, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy (Contract DE-AC 03-76SF00098). Portions of this investigation were completed at the San Diego Supercomputer Center and at the College of Chemistry Graphics Facility at the University of California at Berkeley (funded in part by the National Institutes of Health, Grant No. 8S0RR05651A). R.F. is a recipient of a Kekulē fellowship provided by the Fonds der chemischen Industrie, Germany (1989-1992).

Registry No. 4, 694-87-1; **5**, 1610-51-1; **6**, 58436-35-4; **7**, 60323-52-6; **8**, 20265-84-3; **9**, 93716-74-6; **10**, 93716-73-5; **11**, 143330-04-5; **12**, 4026-23-7; **13**, 24447-42-5; **14**, 61960-82-5; **15**, 69038-28-4.

The Effect of Basis Set and Electron Correlation on the Predicted Electrostatic Interactions of Peptides

Sarah L. Price,*,† Jamie S. Andrews,‡ Christopher W. Murray,‡ and Roger D. Amos‡

Contribution from the Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK, and University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK. *Received April 6, 1992*

Abstract: The realism of molecular modelling studies of peptides depends on the model of the electrostatic forces, and thus on the quality of the wave function used to derive the atomic charges or multipoles. To establish this dependence, we have studied the electrostatic properties of *N*-acetylalanine *N'*-methylamide ($CH_3CONHCHCH_3CONHCH_3$) calculated from a distributed multipole representation of both SCF and correlated wave functions with a range of respectable basis sets. The electron correlation is included in the wave function at second-order Møller-Plesset theory using a "direct" method which calculates the relaxed electron density. To predict the electrostatic potential on the water-accessible surface of the peptide to within a few kJ mol^{-1} requires a correlated wave function and a large basis set of double-zeta plus polarization quality or better. For a given basis set, the SCF wave function overestimates the electrostatic potential by around 15%, the inclusion of electron correlation producing a consistent change in the electron density. Changing the basis set, within a given ab initio method, also produces significant differences in the electrostatic potential around the peptide and in its electrostatic interaction with water. However, it is observed that the electrostatic potential for this peptide correlates strongly with the total dipole moment of the charge density, as the direction of the dipole moment is almost independent of basis set within each ab initio method. Upon correlating the wave function there is a small, almost constant, change in the direction of the dipole moment of 3.5° ± 0.1°. Thus, the electrostatic potential calculated from smaller (split valence) basis sets can be scaled to give good agreement with more accurate calculations.

I. Introduction

The model for the intermolecular forces is the fundamental input to any computer simulation study of the behavior of molecules, and so it is important that the model potentials used are sufficiently accurate for the simulation to be realistic. The search for more reliable model potentials for biologically important molecules, such as polypeptides, has recently focused on the model for the electrostatic interactions, as the contribution which appears to control "molecular recognition", including hydrogen bonding, provided steric constraints are satisfied. Early force fields represented the electrostatic interaction by empirically fitted atomic point charges;[1] however, since this requires Draconian assumptions about the nature of the charge distribution, such charge models have widely been superseded by models derived from the ab initio charge densities of the molecules.

The electrostatic interaction energy in terms of the charge distribution, $\rho^A(\mathbf{r}_i)$, of molecule A and, $\rho^B(\mathbf{r}_j)$, of molecule B is given by:

$$U_{el}^{AB} = \int \frac{\rho^A(\mathbf{r}_i)\rho^B(\mathbf{r}_j)\, d^3\mathbf{r}_i\, d^3\mathbf{r}_j}{4\pi\epsilon_0|\mathbf{R} + \mathbf{r}_j - \mathbf{r}_i|} \qquad (1)$$

where **A** and **B** are arbitrary chosen origins in the molecules and $\mathbf{R} = \mathbf{B} - \mathbf{A}$. This is costly to evaluate so a simplified representation of the charge density is usually defined which is suitable for use in computer simulations. The Mulliken atomic charges[2] were used widely and, indeed, are still used in some commercial modelling packages, but it has often been reported that these give a poor approximation to the electrostatic potential.[3-6] Many alternatives have been proposed. There is considerable interest in the method of potential derived charges,[3,7] where the atomic charges are fitted directly to the electrostatic potential, as evaluated by integration, at a large grid of points in the region of interest outside the molecule. This will clearly give the most accurate representation of the potential in the grid region which is possible with the atomic charges. However, the residual errors in this fitting process emphasize that an atomic point charge model cannot exactly represent the electrostatic forces between molecules, because it explicitly assumes that the charge distribution around each atomic site is spherical. However, the valence electrons of a molecule are often far from spherically distributed, with features such as lone pair and π electron density being invoked to describe chemical bonding and the orientation dependence of hydrogen bonding. Thus another approach to representing the molecular charge distribution is to represent it in terms of sets of multipoles (charge, dipole,

† University College London.
‡ University Chemical Laboratory.

(1) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319.
(2) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(3) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592.
(4) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.
(5) Price, S. L.; Harrison, R. J.; Guest, M. F. *J. Comput. Chem.* **1989**, *10*, 552.
(6) Price, S. L., Richards, N. G. J. *J. Comp. Aid. Molec. Design.* **1991**, *5*, 41.
(7) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

0002-7863/92/1514-8268$03.00/0

quadrupole, octupole, etc) at many sites within the molecule, often with a site at every atom. These distributed multipole models[8-16] can, in principle, be used to calculate the electrostatic potential outside the molecule to the same accuracy as the wave function, provided sufficient sites and multipoles are used. The higher multipole moments automatically represent the electrostatic effects of features such as lone pair or π electron density, though the resulting anisotropy does complicate the model potential.

Although considerable effort has been invested in improving the accuracy of the representation of the ab inito charge density, little attention has been paid to the effect of the quality of the ab initio wave function on the predicted electrostatic properties. Most calculations have been restricted to SCF wave functions, and basis sets of up to 6-31G* in quality, except for the smallest molecules. However, it is well known from studies on small molecules with high quality basis sets that SCF molecular multipole moments are usually 10%–20% in error (typically being too large) and that the effects of electron correlation are significant. This has been shown in calculations by many authors, for example, in the studies of Sadlej, Diercksen, and co-workers.[17-25] For large polyatomics, the central multipole expansion is completely inappropriate for representing electrostatic interactions because it is only valid outside a sphere which contains all the nuclei,[26] assuming that penetration effects are represented separately. For any large molecule which is not approximately spherical, molecular simulations will sample regions within this sphere. Thus, the key question is whether electron correlation produces a sufficient change in the charge distribution that it has a significant effect on the predicted potential in the region around the molecule sampled by other molecules in a simulation.

This paper seeks to answer this question for polypeptides, by using a recently developed method of producing ab initio wave functions for large molecules which include correlation effects. The method is an extension of the direct SCF method,[27] where the conventional limitation on the size of calculations, namely, the storage of the integrals, is overcome by recalculating the integrals at each iteration. This technique has been used to calculate an SCF wave function[5] for an undecapeptide derivative of the immunosuppressant cyclosporin, $C_{63}H_{113}N_{11}O_{12}$, which required 1000 basis functions for a 3-21G basis set. The method is ideal for large molecules, where many of the integrals need not be calculated because they are negligible, since they involve basis functions on centers which are well separated. It is also possible to use direct methodology with second-order Møller–Plesset theory for both energies[28] and gradients.[29-32] The theory developed for

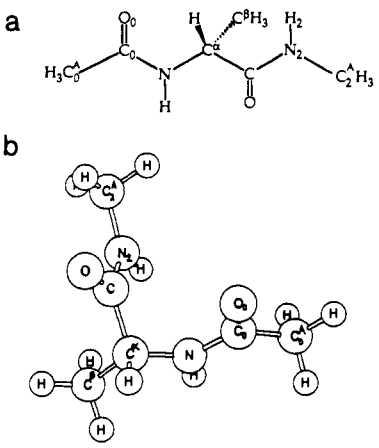

Figure 1. (a) Diagram of *N*-acetylalanine *N'*-methylamide giving notation for atoms. (b) Diagram of *N*-acetylalanine *N'*-methylamide displaying three dimensional conformation (viewed down an axis close to the dipole axis).

gradients also allows the calculation of MP2 corrections to multipole moments and therefore electrostatic potentials. There have been relatively few studies on the effect of correlation on the electrostatic potentials of large molecules. One such study on pyrimidone tautomers is by Les and Adamowicz,[33] and recently Wiberg et al.[34] have studied this effect on several small molecules. In the Appendix the MP2 correlation correction is explained in more detail and is contrasted with that of Les and Adamowicz.[33]

The molecule which was chosen for this study was *N*-acetylalanine *N'*-methylamide, $CH_3CONHCHCH_3CONHCH_3$. This is the smallest model molecule where the central peptide group is subject to the same short-range inductive effects as in a polypeptide. Thus the charge distribution of the peptide should be similar to that found in polypeptides.[35] The electrostatic potential on a grid of points outside this molecule and also the electrostatic interaction with a water molecule were calculated using accurate distributed multipole representations of a wide range of wave functions calculated using different basis sets, at both the SCF and MP2 level. We have used direct MP2 gradient codes to calculate MP2 corrections with up to 220 basis functions. The results show that the quality of the basis set and the inclusion of electron correlation have an important effect on the absolute values of the electrostatic potential. However, the size of calculation which is necessary to predict the electrostatic potential close to the molecule to within a few kJ mol^{-1} is large. Hence, we also assess the accuracy of various common assumptions, such as the reliability of the relative electrostatic potentials, or of scaling the electrostatic model.

II. Methods

The calculations were carried out on *N*-acetylalanine *N'*-methylamide in a fixed geometry, defined using the standard bond lengths and angles of the AMBER force field,[36] and the torsion angles $\phi = -57°$, $\psi = -47°$ of α-poly-L-alanine[37] (see Figure 1). Thus, the charge distribution of the central peptide group should be a fair approximation to that found for an alanine residue in an α-helix.

The basis sets used in this work were STO-3G, 3-21G, 3-21G*, 3-21G**, 4-31G, 4-31G*, 4-31G**, 6-31G, 6-31G*, and 6-31G** as described by Pople,[38] the DZ basis set of Dunning,[39] a DZP

(8) Rein, R. *Adv. Quant. Chem.* **1973**, *7*, 335.
(9) Mezei, M.; Campbell, E. S. *Theor. Chim. Acta.* **1977**, *43*, 227.
(10) Pullman, A.; Perahia, D. *Theor. Chim. Acta* **1978**, *48*, 29.
(11) Lavery, R.; Etchebest, C.; Pullman, A. *Chem. Phys. Lett.* **1982**, *85*, 266.
(12) Rico, J. F.; Alvârez-Collado, J. R.; Paniagua, M. *Mol. Phys.* **1985**, *56*, 1145.
(13) Cooper, D. L.; Stutchbury, N. C. J. *Chem. Phys. Lett.* **1985**, *120*, 167.
(14) Sokalski, W. A.; Sawaryn, A. *J. Chem. Phys.* **1987**, *87*, 526.
(15) Vigné-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934.
(16) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233. Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.
(17) Diercksen, G. H. F.; Sadlej, A. J. *Chem. Phys.* **1981**, *61*, 293.
(18) Diercksen, G. H. F.; Kellö, V.; Sadlej, A. J. *J. Chem. Phys.* **1983**, *79*, 2918.
(19) Diercksen, G. H. F.; Kellö, V.; Sadlej, A. J. *Mol. Phys.* **1983**, *49*, 711.
(20) Diercksen, G. H. F.; Sadlej, A. J. *Chem. Phys.* **1985**, *96*, 17.
(21) Diercksen, G. H. F.; Sadlej, A. J. *Chem. Phys. Lett.* **1985**, *114*, 187.
(22) Diercksen, G. H. F.; Sadlej, A. J. *Mol. Phys.* **1986**, *57*, 509.
(23) Kellö, V.; Roos, B. O.; Sadlej, A. J. *Theor. Chim. Acta* **1988**, *74*, 185.
(24) Kellö, V.; Noga, J.; Diercksen, G. H. F.; Sadlej, A. J. *Chem. Phys. Lett.* **1988**, *152*, 387.
(25) Diercksen, G. H. F.; Sadlej, A. J. *Chem. Phys. Lett.* **1988**, *153*, 93.
(26) Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 3325.
(27) Almlöf, J.; Faegri, K.; Korsell, K. *J. Comput. Chem.* **1982**, *3*, 385.
(28) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.
(29) Sæbø, S.; Almlöf, J. *Chem. Phys. Lett.* **1989**, *154*, 83.
(30) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275.
(31) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 281.
(32) Sæbø, S. *Int. J. Quantum Chem.* **1990**, *38*, 641.
(33) Les, A.; Adamowicz, L. *Chem. Phys.* **1991**, *153*, 409.
(34) Wiberg, K. B.; Hadad, C. M.; Lepage, T. J.; Breneman, C. M.; Frisch, M. J. *J. Phys. Chem.* **1992**, *96*, 671.
(35) Faerman, C. H.; Price, S. L. *J. Am. Chem. Soc.* **1990**, *112*, 4915.
(36) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(37) Arnott, S.; Dover, S. D. *J. Mol. Biol.* **1967**, *30*, 209.
(38) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio Molecular Orbital Theory*, Wiley-Interscience: New York, 1986.
(39) Dunning, T. H. *J. Chem. Phys.* **1970**, *53*, 2823.

Table I. Basis Set Dependence of the Total Energy and Dipole Moment of *N*-Acetylalanine *N'*-Methylamide[a]

basis	SCF energy ($10^3 E_h$)	SCF dipole magnitude μ (ea_0)	SCF dipole orientation		MP2 energy (E_h)	MP2 dipole magnitude μ (ea_0)	MP2 dipole orientation	
			a (deg)	*b* (deg)			*a* (deg)	*b* (deg)
STO-3G	−0.486 55	2.07	19.8	19.5				
3-21G	−0.490 10	2.92	16.6	16.2	−1.007 51	2.45	20.4	20.0
3-21G*	−0.490 48	2.67	17.1	16.7	−1.477 15	2.32	20.7	20.2
3-21G**	−0.490 51	2.68	16.9	16.5				
4-31G	−0.492 12	3.14	16.7	16.2	−1.022 85	2.68	20.2	19.7
4-31G*	−0.492 37	2.94	17.0	16.5				
4-31G**	−0.492 39	2.95	16.8	16.3				
6-31G	−0.492 62	3.17	16.6	16.2	−1.011 84	2.71	20.2	19.7
6-31G*	−0.492 84	2.96	17.0	16.5	−1.494 01	2.61	20.3	19.8
6-31G**	−0.492 87	2.96	16.9	16.4	−1.571 37	2.56	20.3	19.8
DZ	−0.492 70	3.32	16.5	16.0	−1.091 71	2.88	19.9	19.4
DZP	−0.492 96	3.07	17.0	16.5	−1.653 68	2.67	20.5	19.9
5s4p2d	−0.493 03	3.01	16.8	16.2				

[a] The dipole moment is at an angle $a°$ to the α-helix axis, at an angle $b°$ to the plane containing this axis and C_α, so that it points approximately down the helix, with a slight tilt toward C. The direction and position of the helix axis were defined by matching the C_α position, C_α–C bond direction, and $C_\alpha CO$ plane to those of the cylindrical polar coordinates of polyalanine.[37]

basis (DZ plus one set of polarization functions: C(0.8), N(0.8), O(0.9), H(1.0)), and a 5s4p2d basis (the triple-zeta (ζ) set of Dunning[40] with two sets of polarization functions: C(1.2,0.4), N(1.35,0.45), O(1.35,0.45), H(1.5,0.5)).

In order to confirm the convergence of our results for *N*-acetylalanine *N'*-methylamide as a function of basis set, we performed similar calculations on formamide, in the corresponding standard geometry,[36] with the basis sets defined above and with the larger 8s6p3d basis set: the 8s6p set is a contraction of the van Duijneveldt 12s9p primitive set[41] with polarization functions C(1.8,0.6,0.2), N(2.4,0.8,0.3), and O(2.7,0.9,0.3). The hydrogen basis at this level is a 6s contraction of the van Duijneveldt 10s basis, with polarization functions having exponents 1.8, 0.6, and 0.2.

A distributed multipole analysis (DMA)[16] of each wave function was obtained, using the option within CADPAC.[42] The DMA uses the density matrix of the wave function, expressed in terms of the Gaussian primitives which comprise the basis set. Each term can be represented exactly by a finite set of multipoles at the overlap center corresponding to the two Gaussian primitives involved. If this overlap center is not one of the DMA expansion sites, then these multipoles are represented at the nearest expansion site, thus optimizing the convergence of the resultant multipole series at each site. The wave functions were represented by a set of multipoles, up to and including the hexadecapole moment, on each atomic site, requiring a total of 25 multipole components for each of the 22 atoms. This is more sites than would be normally appropriate for modelling peptides, as it has been shown[35] that not having any multipoles on nonpolar hydrogen atoms, but representing their effects by the multipoles on the bonded carbon atom, produces negligible errors in the potential at 1 Å from the van der Waals surface. However, for the purposes of this study, it is important that the representation of the wave function does not introduce any errors in the predicted potential. The DMA representation does not include the effects of overlap of the two molecular charge distributions on the electrostatic energy, as such penetration effects depend exponentially on separation, and therefore are more appropriately represented separately, or absorbed into the other short range forces which produce the repulsive wall.[26]

The electrostatic potential, field and electrostatic interaction energies were calculated from the DMAs using all terms in the multipole expansion up to R^{-5}, within the program ORIENT.[43] Three different criteria were used in comparing the electrostatic properties. Firstly, we compared the value of the predicted electrostatic potential at a few fixed points, including the positions of the atoms which would be hydrogen bonded to the alanine residue in an α-helix. Secondly, we compared the potential on a grid of 3830 points, defined to be 1.4 Å from the van der Waals surface of the molecule, spaced by approximately 0.3 Å. The van der Waals surface of the molecule was defined by the Pauling radii: 1.5 Å for N, 1.4 Å for O, and 2.0 Å for C atoms. There was no explicit hydrogen van der Waals radius, as the hydrogen atom is included in a "united atom" methyl radius, and polar hydrogen atoms effectively have no radius when involved in hydrogen bonding. The use of a grid of points 1.4 Å from the van der Waals surface approximates the water-accessible surface of the molecule and is as close as any atom is likely to approach the molecule, except hydrogen bonding protons and some small ions. The results suggest that the choice of grid is unlikely to affect the qualitative conclusions of the study.

The values of the potential and field quantify the interaction of the molecule with a point charge and point dipole. However, in most molecular modelling, it is the interaction with other molecules, which have a finite size and varying charge distribution, that is important, particularly the prediction of the optimum relative orientation of the two molecules. Thus, for the third criterion, we have predicted the water binding sites around *N*-acetylalanine *N'*-methylamide, by optimizing the electrostatic energy of the peptide with water, using a DMA of water calculated with the same quality wave function. These optimizations were constrained to sterically accessible orientations, defined using a pseudo-hard sphere repulsive potential, with the van der Waals radii given above. The comparison of the positions and magnitudes of the minima in the electrostatic interaction energy is a useful indicator of the similarity of the electrostatic properties. The belief that minimizing the electrostatic energy will predict the site at which one water molecule would bind to the peptide, in isolation, follows from the success of such a model in predicting the geometries of van der Waals dimers,[44] and the orientational preference of N—H···O=C hydrogen bonds.[45] This success can be attributed to the electrostatic contribution generally mirroring the orientation dependence of the total energy in hydrogen bonding interactions,[46,47] provided that it is calculated from a DMA or another representation which includes the electrostatic effects of lone pairs and π electrons.

III. Results

A. The Monomer Energies and Total Dipole Moments. Table I presents the total SCF and MP2 correlation energies for *N*-

(40) Dunning, T. H. *J. Chem. Phys.* **1971**, *55*, 716.
(41) van Duijneveldt, F. B. IBM Research Report, RJ945, 1971.
(42) CADPAC5: The Cambridge Analytical Derivative Package, Issue 5, Cambridge, 1992. A suite of quantum chemistry programs written by R. D. Amos, I. L. Alberts, J. S. Andrews, S. M. Colwell, N. C. Handy, D. Jayatilaka, P. J. Knowles, R. Kobayashi, K. E. Laidig, P. E. Maslen, C. W. Murray, J. E. Rice, E. D. Simandiras, A. J. Stone, and M-D. Su.
(43) Stone, A. J. ORIENT, Version 2: a program for calculating electrostatic interactions between molecules, (University of Cambridge, 1990).
(44) Buckingham, A. D.; Fowler, P. W. *Can J. Chem.* **1985** *63*, 2018.
(45) Mitchell, J. B. O.; Price, S. L. *Chem. Phys. Lett.* **1989**, *154*, 267.
(46) Hurst, G. J. B.; Fowler, P. W.; Stone, A. J.; Buckingham, A. D. *Int. J. Quantum Chem.* **1986**, *29*, 1223.
(47) Mitchell, J. B. O.; Price, S. L. *J. Comput. Chem.* **1990**, *11*, 1217.

Table II. Comparison of ab Initio Electrostatic Properties of Formamide

basis	6-31G**	DZP	5s4p2d	8s6p3d
(a) SCF Wave Functions				
energy (E_h)	−168.937 21	−168.973 77	−169.001 02	−169.009 53
dipole (ea_0)	1.727	1.799	1.779	1.759
dipole angle θ^a (deg)	41.6	42.0	41.8	41.7
quadrupole (ea_0^2)	4.151	4.337	4.337	4.359
electrostatic potential at hydrogen bond positions[b]				
H H-bond V (kJ mol^{-1})	112	116	115	114
O H-bond (1) V (kJ mol^{-1})	−169	−176	−175	−173
O H-bond (2) V (kJ mol^{-1})	−186	−193	−190	−187
electrostatic potential on water-accessible surface				
rms V (kJ mol^{-1})	57.3	59.7	59.0	58.3
max V (kJ mol^{-1})	111.5	114.7	113.7	112.8
min V (kJ mol^{-1})	−99.8	−103.1	−102.1	−100.7
ΔV = V(basis, SCF) − V(8s6p3d,MP2)				
rms ΔV (kJ mol^{-1})	5.9	8.1	7.4	6.9
largest positive ΔV (kJ mol^{-1})	9.8	14.1	12.3	11.2
largest negative ΔV (kJ mol^{-1})	−11.3	−14.4	−13.6	−12.3
ΔV = V(basis,SCF) − V(8s6p3d,SCF)				
rms ΔV (kJ mol^{-1})	1.4	1.4	0.7	
largest positive ΔV (kJ mol^{-1})	2.6	3.1	1.5	
largest negative ΔV (kJ mol^{-1})	−2.9	−2.5	−1.4	
(b) MP2 Correlated Wave Functions				
MP2 energy (E_h)	−0.493 66	−0.520 70	−0.637 67	−0.708 442
dipole (ea_0)	1.531	1.616	1.603	1.580
dipole angle θ^a (deg)	37.3	38.5	38.6	38.5
quadrupole (ea_0^2)	3.796	4.007	4.034	4.062
electrostatic potential at hydrogen bond positions[b]				
H H-bond V (kJ mol^{-1})	104	108	107	106
O H-bond (1) V (kJ mol^{-1})	−145	−154	−155	−154
O H-bond (2) V (kJ mol^{-1})	−163	−172	−171	−167
electrostatic potential V on water-accessible surface				
rms V (kJ mol^{-1})	50.7	53.6	53.2	52.5
max V (kJ mol^{-1})	102.3	106.6	106.2	105.4
min V (kJ mol^{-1})	−86.3	−90.6	−90.5	−88.9
ΔV = V(basis,MP2) − V(8s6p3d,MP2)				
rms ΔV (kJ mol^{-1})	2.5	1.4	0.8	
largest positive ΔV (kJ mol^{-1})	5.2	2.9	1.7	
largest negative ΔV (kJ mol^{-1})	−5.7	−2.1	−1.6	
ΔV = V(basis,SCF) − V(basis,MP2)				
rms ΔV (kJ mol^{-1})	8.0	7.2	6.8	6.9
largest positive ΔV (kJ mol^{-1})	13.4	11.4	10.8	11.2
largest negative ΔV (kJ mol^{-1})	−14.1	−13.2	−12.2	−12.3

[a] The dipole moment is at θ to the CN bond vector, on the opposite side to the O atom. [b] The positions of two likely sites of protons hydrogen bonded to the O atom, and of O atoms hydrogen bonded to the trans N–H group, were defined by an assuming that the hydrogen bond would be coplanar with the molecule, with an O···H hydrogen bond length of 1.88 Å, a NH···O angle of 180°, and a CO···H angle of 138°. Of the two sites for hydrogen bonding to O, site (1) is closer to the N atom.

acetylalanine *N'*-methylamide, for the range of calculations. We have calculated the MP2 correction for several of the smaller basis sets. However, it should be noted that, in general, it is unadvisable to use small basis sets with a correlated method, and that 6-31G* is probably the smallest basis that ought to be used. The total SCF energy decreases by $3E_h$ (E_h = energy in hartrees) in going from a 3-21G to a DZP basis, which is approximately twice the maximum MP2 correlation energy. The total dipole moment is also tabulated as a simple measure of the distribution of charge in the model. The magnitude of the dipole moment is very sensitive to basis set, with values differing by up to 25% from the dipole moment of the correlated DZP wave function. The neglect of correlation, for a given basis, overestimates the dipole moment by up to 20%. The addition of polarization functions has relatively little effect on the SCF energy, but increases the correlation energy by around 50%, and decreases the dipole moment by around 7%. The addition of polarization functions on the H atoms in the split-valence basis sets has a negligible effect at the SCF level. There is a general increase in the dipole moment with an improved description of the s and p orbitals. The decrease in the magnitude of the dipole moment with the addition of polarization functions or correlation implies that the molecular charge distribution becomes more compact (given the opportunity).

A rather less predictable result is that the orientation of the dipole moment is remarkably constant as the basis set is varied using the same ab initio method. The two angles required to define the direction of the dipole moment vary by less than a degree for all the correlated wave functions and also for the set of SCF wave functions (excluding the STO-3G basis set). Calculating the scalar product between the SCF dipole and MP2 dipole, within the same basis, shows that the dipole moment changes direction by 3.5° ± 0.1° on correlating a given wave function.

B. Convergence of Calculations. In order to check the convergence of our results, some further large basis set calculations were made on the smaller formamide molecule. The results are given in Table II. In practical terms the definition of convergence depends upon the property being considered, and the use to which it will be put. In the case of the electrostatic potential, changes which correspond to differences in interaction energies of about 1 kcal/mol (≈4 kJ mol^{-1}) can be ignored. From the formamide results it is seen that the larger basis sets (DZP, 5s4p2d,8s6p3d) are converged at both SCF and MP2 levels, with the possible exception of the potential close to the hydrogen bonding positions. There are of course higher-order correlation effects beyond MP2. However, for closed-shell molecules well described by SCF methods, the MP2 correction will be the bulk of the correlation correction (see, for example, the work of Sadlej and Dierckxsen[17-25]), and the remaining corrections will be small (of the order of the remaining basis set errors). One exception to this might be the potential close to a triply-bonded group (e.g., CN), but such groups do not occur in peptides. Thus the "in vacuo" electrostatic properties calculated from the MP2 DZP wave

Table III. Electrostatic Properties of *N*-Acetylalanine *N'*-Methylamide 1.4 Å from the van der Waals Surface

basis	rms V (kJ mol^{-1})	max V (kJ mol^{-1})	contact atom[a]	min V (kJ mol^{-1})	rms $\|E\|$ (kJ mol^{-1} Å^{-1})	max $\|E\|$[b] (kJ mol^{-1} Å^{-1})
			(a) SCF Wave Functions			
STO-3G	41.3	97.0	N_2*	−112.5	12.7	32.0
3-21G	58.4	127.8	N	−164.3	18.2	43.4
3-21G*	53.7	123.8	N_2*	−147.2	16.7	42.3
3-21G**	53.7	123.4	N_2*	−147.2	16.7	42.7
4-31G	62.9	134.4	N	−177.5	19.6	44.9
4-31G*	59.0	130.3	N_2*	−163.2	18.3	44.3
4-31G**	59.0	129.9	N	−163.4	18.3	44.7
6-31G	63.4	135.4	N	−178.7	19.8	45.3
6-31G*	59.3	131.0	N_2*	−164.2	18.4	44.7
6-31G**	59.3	130.6	N*	−164.4	18.4	45.0
DZ	66.3	140.1	N*	−185.5	20.6	45.2
DZP	61.3	133.7	N_2*	−169.7	19.1	43.9
5s4p2d	60.3	131.5	N_2*	−166.3	18.8	43.3
			(b) MP2 Correlated Wave Functions			
3-21G	49.6	116.2	N_2*	−136.2	15.5	40.3
3-21G*	47.1	114.9	N_2*	−126.6	14.7	40.0
4-31G	54.0	121.8	N_2*	−149.7	16.8	41.7
6-31G	54.6	123.2	N_2*	−151.3	17.0	42.2
6-31G*	52.6	122.2	N_2*	−143.9	16.4	42.1
6-31G**	51.7	120.1	N_2*	−141.6	16.1	41.7
DZ	58.0	129.8	N_2*	−160.7	18.1	42.4
DZP	54.0	123.9	N_2*	−147.9	16.8	41.1

[a] An asterisk indicates that the maximum was close to the intersection of the surface of the given atom with that of another atom. The minimum value of V always occurs at the same point on the surface of the oxygen atom O, near its intersection with the surface of O_0. [b] The maximum value of the field magnitude $|E|$ always occurs at the same point on the N atom surface.

Table IV. Errors in the Electrostatic Properties of *N*-Acetylalanine *N'*-Methylamide 1.4 Å from the van der Waals Surface

basis	rms ΔV (kJ mol^{-1})	largest positive ΔV (kJ mol^{-1})	contact atom[a]	largest negative ΔV (kJ mol^{-1})	contact atom	rms scaled ΔV[b] (kJ mol^{-1})
		(a) SCF Wave Functions: $\Delta V = V$(basis,SCF) − V(DZP,MP2)				
STO-3G	12.9	35.7	O	−27.7	N_2*	2.6
3-21G	6.0	9.0	C_0^A	−17.4	O*	3.9
3-21G*	2.9	5.3	O_0*	−5.9	C_2^A	2.9
3-21G**	3.1	5.3	O_0*	−6.2	C_2^A	3.1
4-31G	9.9	15.3	C_0^A*	−29.9	O*	4.0
4-31G*	5.9	10.6	C_0^A*	−16.0	O*	3.0
4-31G**	6.0	10.7	C_0^A*	−16.2	O*	3.2
6-31G	10.3	16.2	C_0^A*	−30.9	O*	3.9
6-31G*	6.2	11.4	C_0^A*	−16.9	O*	3.0
6-31G**	6.3	11.3	C_0^A*	−17.0	O*	3.1
DZ	13.0	20.3	C_0^A*	−37.6	O*	3.8
DZP	8.0	13.4	C_0^A*	−21.9	O*	3.0
5s4p2d	7.1	10.9	C_0^A*	−18.5	O*	3.2
		(b) MP2 Correlated Wave Functions: $\Delta V = V$(basis,MP2) − V(DZP,MP2)				
3-21G	4.6	13.2	C*	−7.7	N_2*	1.6
3-21G*	7.2	21.4	O_0*	−9.3	N_2*	2.1
4-31G	1.4	4.4	C	−4.2	O	1.4
6-31G	1.5	3.7	C	−5.0	O	1.3
6-31G*	1.5	4.3	O	−2.3	C_0^A*	0.7
6-31G**	2.3	6.5	O	−4.0	C^β	0.7
DZ	4.3	6.8	N	−12.8	O*	1.2

[a] An asterisk indicates that point was close to intersection of surface of given atom with another atom. [b] Rms value of V(DZP,MP2) − V(basis, method)[μ(DZP,MP2)]/[μ(basis,method)].

function for *N*-acetylalanine *N'*-methylamide are probably as accurate as will ever be required for simulation work.

C. The Electrostatic Potential Outside *N*-Acetylalanine *N'*-Methylamide. The predicted electrostatic potential at three points close to the molecule, including hydrogen bonding sites, have been plotted in Figure 2 against the magnitude of the total dipole moment for the corresponding wave function. The variation in predicted potential with wave function is clearly a matter for concern, with the commonly used 6-31G* SCF wave function giving an error, relative to the MP2 DZP result, of more than 20 kJ mol^{-1} near the oxygen atom (VO). The minimal basis set (STO-3G) predicts potential values that are far too small in magnitude. Otherwise, the SCF results generally overestimate the potential magnitudes relative to the correlated wave functions. However, the correlation between the predicted potential and the magnitude of the total dipole moment is surprisingly good, with correlation coefficients for each ab initio method of 0.97 or better at the two hydrogen bonding points, and only slightly worse at the point outside C_α.

The results at the chosen points are consistent with the analysis of the potential on the water-accessible surface (Table III). The differences in the root-mean-square (rms) values of the potential and field are quantitatively significant. However, the overall picture of the electrostatic potential around the molecule, when viewed on a graphics screen, is qualitatively very similar for different wave functions, with a smooth variation in the potential from a maximum in the region over the two protons of the NH groups, to a minimum over the two CO groups.[48] The position of the maximum field and potential minimum was independent

(48) Price, S. L.; Stone, A. J. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 1755.

Table V. Effect of Basis Set on SCF Calculations of the Electrostatic Properties of *N*-Acetylalanine *N'*-Methylamide 1.4 Å from the van der Waals Surface

basis	rms ΔV (kJ mol^{-1})	largest positive ΔV (kJ mol^{-1})	contact atom[a]	largest negative ΔV (kJ mol^{-1})	contact atom[a]	rms scaled ΔV (kJ mol^{-1})
		ΔV = V(basis,SCF) − V(DZP,SCF)				
STO-3G	20.3	57.1	O*	−39.6	N	3.9
3-21G	3.4	9.1	O	−8.3	C_2^A	1.9
3-21G*	7.8	22.5	O_0*	−11.1	N*	1.9
3-21G**	7.8	22.5	O_0*	−11.1	N*	1.8
4-31G	2.4	3.4	C	−8.1	O_0	1.8
4-31G*	2.5	6.7	O	−3.6	N_2*	0.8
4-31G**	2.5	6.6	O	−4.0	C^β	0.8
6-31G	2.7	3.1	C_0^A	−9.1	O_0	1.7
6-31G*	2.1	5.6	O	−2.8	C^β	0.7
6-31G**	2.1	5.5	O	−3.4	C^β	0.7
DZ	5.1	7.4	N	−15.9	O*	1.4
5s4p2d	1.2	3.4	O*	−3.3	N*	0.5

[a] An asterisk denotes that error occurred at a point on the surface of the given atom close to the intersection with that of other atoms. [b] Rms value of V(DZP,SCF) − V(basis,SCF)[μ(DZP,SCF)]/[μ(basis,SCF)].

Table VI. Effect of Correlation on Calculations of the Electrostatic Properties of *N*-Acetylalanine *N'*-Methylamide 1.4 Å from the van der Waals Surface[a]

basis	rms ΔV (kJ mol^{-1})	largest positive ΔV (kJ mol^{-1})	largest negative ΔV (kJ mol^{-1})	rms $\lvert\Delta E\rvert$ (kJ mol^{-1} Å^{-1})	largest $\lvert\Delta E\rvert$ (kJ mol^{-1} Å^{-1})
		ΔV = V(basis,SCF) − V(basis,MP2)			
3-21G	9.6	14.3	−28.1	3.1	6.7
3-21G*	7.3	11.4	−20.5	2.3	4.9
4-31G	9.5	14.6	−27.9	3.1	6.7
6-31G	9.5	14.6	−27.6	3.0	6.6
6-31G*	7.3	11.8	−20.4	2.3	4.9
6-31G**	8.2	13.6	−22.8	2.6	5.5
DZ	8.9	14.5	−25.2	2.8	5.9
DZP	8.0	13.4	−21.9	2.5	5.1

[a] The largest positive errors were all at the same point on the surface of C_0^A, near the intersection with the surface of NH. The largest errors in $|E|$ and the largest negative errors were always close to the intersection of the water-accessible surfaces of the O_0 and O.

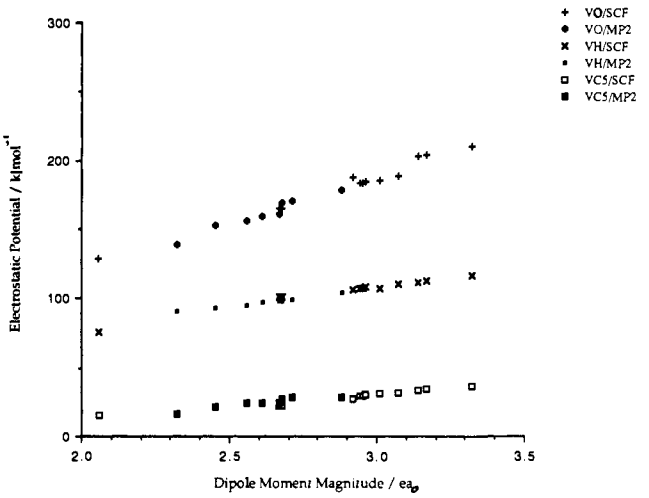

Figure 2. The electrostatic potential at three points outside *N*-acetylalanine *N'*-methylamide as a function of the total dipole moment magnitude, for a range of SCF and MP2 wave functions. The point VO was 2.06 Å from O, and VH 2.06 Å from the amide H, at orientations[47] typical of those of the hydrogen bond acceptor or donor in an α helix.[64] The point VC was 5 Å from the helix axis, on the radial vector through C_α and is therefore 2.71 Å from C_α.

of wave function, and the position of the potential maximum was usually close to the intersection of the van der Waals surfaces of the two N atoms.

An analysis of the errors in the predicted potential, compared with that calculated from the MP2 DZP wave function (Table IV), shows that rms errors of up to 10 kJ mol^{-1} can arise within the range of respectable basis sets, and there is some variation in the position of the maximum and minimum errors. The various competing factors which affect the total dipole moment contribute to the potential differences, and it appears that an SCF calculation with a 3-21G* basis set predicts the potential in good agreement with the definitive values. This basis set gives quite severe errors relative to the SCF limit (Table V), but, in this case, these approximately cancel with the errors arising from the neglect of correlation (Table VI). This cancellation will not necessarily occur for other molecules. Varying the basis set within the SCF or correlated method does not give a simple trend in the errors, and the positions of the worst errors are quite variable. However, correlating a wave function within a given basis set (Table VI) produces a remarkably consistent change in the electrostatic potential with a rms difference of 8.5 ± 0.9 kJ mol^{-1} in the value of the potential.

D. Scaling Corrections. The success of the correlation between the potential at certain points and the magnitude of the total dipole moment in Figure 2 suggests that the predictions of the less accurate wave functions could be improved by scaling the potential.

$$V_{\text{scaled}} = V(\text{basis,method})[\mu(\text{DZP,MP2})/\mu(\text{basis,method})] \quad (2)$$

Scaling the potential from the various correlated wave functions gives quite good predictions of the definitive potential, reducing the rms error to 2 kJ mol^{-1} or less (Table IV). Similarly, the various SCF wave functions give close agreement on the electrostatic potential when scaled by the corresponding total dipole moments (Table V). However, the various SCF wave functions are unable to predict the definitive MP2 DZP potential any better than the 3-21G* SCF wave function; there is a remarkably constant residual rms error of 3 to 4 kJ mol^{-1} (Table IV), suggesting a consistent difference between the SCF and correlated wave functions. (The exception to this is the SCF STO-3G charge distribution, where the total dipole direction is closer to that of the correlated wave functions than the other SCF wave functions. The corresponding potential can be scaled into closer agreement with the MP2 DZP potential than the SCF DZP potential.) The total dipole direction is clearly a key indicator of the accuracy of scaling the potential.

Scaling the potential will also be successful at very long range, where the central multipole expansion is converged, and only the

total dipole moment contributes significantly to the electrostatic potential. However, the potential only scales with the magnitude of the dipole moment for both the SCF and correlated sets of wave functions because the direction of the dipole moment is so insensitive to the basis set. The approximately constant difference in the direction of the dipole moment between SCF and correlated wave functions will ensure that scaling by dipole magnitudes will not bring the potential calculated from SCF wave functions into exact agreement with that predicted that correlated wave functions. Thus, our results much closer to the molecule are unlikely to be qualitatively changed by varying the position of the surface grid points.

In summary, the values of the electrostatic potential predicted by this range of respectable ab initio wave functions show a quantitative variation which will be significant for many applications. However, these errors can be significantly reduced by scaling the predicted potential by the ratio of the total dipole moment of the wave function to the accurate dipole moment. Nevertheless, the scaling of SCF values for the potential cannot give exact agreement with values derived from correlated wave functions, leaving a fairly constant residual rms error in the potential. The consistency of the change in the direction of the total dipole moment and the predicted potential outside the molecule, with the correlation of the wave functions, suggest that correlation has a fairly systematic, non-uniform, effect on the charge distribution of a molecule.

E. The Predicted Water Binding Sites. There are five distinct minima in the electrostatic interaction of water with *N*-acetylalanine *N'*-methylamide in the α helical conformation. The lowest minimum occurs for a water molecule hydrogen bonded to both carbonyl oxygens. There are four local minima where the water molecule has only one hydrogen bond, to each oxygen and to each NH group. These are plausible water binding sites for the isolated molecule. The exact positions of these minima were almost independent of the quality of the wave function used. The positions of the water oxygen atoms hydrogen bonded to NH were within 0.1 Å of the position predicted using MP2 DZP wave functions, in all cases. The positions of water when acting as the proton donor were slightly less consistent, but nevertheless varied by less than 0.2 Å, except in the case of the STO-3G basis, where the global minimum had shifted by 0.3 Å. Although the separation of the water oxygen and the peptide O and NH groups is dictated by the pseudo-hard sphere radii, there is still sufficient orientational freedom that the agreement on the position of the minima is encouraging.

Unfortunately, as Figure 3 demonstrates, the electrostatic energies at these sites are very dependent on the choice of wave function. The inclusion of correlation or the addition of polarization functions can often result in changes of around 10 kJ mol^{-1} in the electrostatic binding energy. The relative qualitative trend of the doubly hydrogen bonded site being the lowest in energy, then the sites hydrogen bonded to one O, and then those bonded to NH being the least favorable is maintained. However, the detailed results, such as the order of energies of water singly hydrogen bonded to O and to O_0 change with wave function.

There is no obvious method of predicting the MP2 DZP water binding energies more accurately from cheaper monomer calculations. The SCF 3-21G* calculation, which gave quite good agreement with the MP2 DZP calculation for the electrostatic potential, does not do particularly well at predicting the electrostatic interaction energies, as the SCF 3-21G* and MP2 DZP dipole moments for water differ by 7%. The variation in the energy differences between the minima show that scaling using the total dipole moments of both molecules could not be very accurate.

IV. Conclusions

Typically electron correlation significantly reduces the strength of the electrostatic forces between molecules, and so electrostatic models based on SCF wave functions with "near-Hartree–Fock limit" basis sets, such as the popular 6-31G* basis, will overestimate the electrostatic potential around the molecule "in vacuo" by around 10%. However, our results show that correlation does not have a uniform effect on the SCF charge distribution, such as would give rise to a constant scaling.

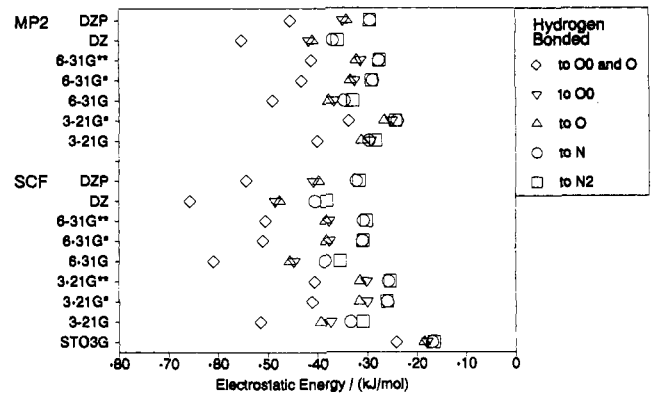

Figure 3. Minima in the electrostatic interaction between *N*-acetylamine *N'*-methylamide and water, as a function of the quality of the wave function. The water molecule has a bond length of 0.96 Å and bond angle of 104.5°. The five minima, found in essentially the same position for all molecular charge distributions, are shown according to the hydrogen bonding in the complex. Three calculations, SCF 6-31G*, SCF and MP2 DZP, also gave a local minimum corresponding to a doubly hydrogen bonded structure like the global minimum, but with the plane of the water molecule closer to the plane of the CO groups. However, the existence of slight secondary minima so close to global minima is unlikely to be significant.

The introduction of correlation into the wave function appears to produce a small redistribution of the charge which alters the direction of the dipole moment, and the predicted potential, in a manner that is fairly independent of basis set. This part of the correlation effect is approximately given by the errors from the scaled SCF wave function

$$\Delta V = V(\mathrm{DZP,MP2}) - V(\mathrm{DZP,SCF})\frac{\mu(\mathrm{DZP,MP2})}{\mu(\mathrm{DZP,SCF})} \quad (3)$$

These differences run from a maximum of 7.6 kJ mol^{-1} above the N_2 proton to a minimum of −5.5 kJ mol^{-1} on the C_0^A methyl group at the other end of the molecule, in a progressive way, rather as if the main effect of correlation had been to shift charge along the peptide bonds of the molecule. This effect is not particularly large for this molecule, but it is sufficient to limit the accuracy with which the potential can be predicted by scaling uncorrelated wave functions.

Other workers have recently noted that electron correlation has an effect which is related to the functional groups and structure of the molecule. The electrostatic potential outside 2(1*H*)-pyrimidone in the region involved in its tautomerism is significantly altered by the inclusion of electron correlation, though the potential in other regions outside the van der Waals surface, and around 2-hydroxypyrimidine, is altered by less than 10%.[33] This probably reflects a large correlation correction in the carbon–oxygen double bond. Studies[49] on molecules with only one or two nonhydrogen atoms revealed that electron correlation made the most difference to the molecular electrostatic potential near the nucleii, with the potential outside the van der Waals surface being relatively unaffected. However, this conclusion may result from the small size and high symmetry of the molecules, and is relative to the changes in the potential inside the molecule, which were the focus of the study. Thus it seems probable that different functional groups may be affected differently by correlation, and more molecules need to be studied before we are sure of its importance for the predicted electrostatic properties inside and outside the molecule.

The question as to the quality of ab initio wave function that is necessary in order to derive reliable electrostatic models for molecular modelling is clearly a difficult one. As with all mo-

(49) Luque, F. J.; Orozco, M.; Illas, F.; Rubio, J. *J. Am. Chem. Soc.* **1991**, *113*, 5203.

lecular modelling work, it will be a matter of pragmatic compromise. Within the range of reasonable basis sets (i.e., excluding STO-3G), the difference in the quality of the wave function does not change the qualitative picture of the electrostatic potential around the molecule. The positions of the extremes of the potential, and the water binding sites, are well predicted. However, there are quantitative differences in the electrostatic potential and energies which are chemically significant. However, these errors will often be smaller than those involved in using a poor representation of the charge distribution, such as Mulliken charges, or in assuming that the charge distribution is independent of conformation.[48]

A common method of improving predicted electrostatic energies is to scale the atomic charges, derived from ab initio or semiempirical wave functions, to give agreement with results of superior calculations.[50-52] There have been various studies to establish the validity of such scalings, particularly comparing semiempirical charges with those derived from reasonable (e.g., 6-31G*) SCF calculations, in an effort to improve the charge models used for larger molecules where even quality SCF calculations are prohibitive. It is the predicted electrostatic potential that is important for the success of the molecular modelling and so we should compare potentials rather than charges. (Although an exact scaling of the charges will lead to an exact scaling of the potential, an approximate scaling of the charges can lead to a significantly worse, or better, scaling of the predicted potential.[53]) Our results show that scaling the potential by the ratio of dipole moment magnitudes is surprisingly successful for this particular structure. The extensive study of scaling for different atomic charge representations for 29 small molecules[53] showed that the correlation between the potential derived charges, and also between the corresponding electrostatic potential outside the molecule, was very high for 3-21G and better wave functions compared to a standard SCF 6-31G* values. This suggests that our conclusion about the scaling of the potential with reasonable basis sets may be applicable to a wide range of molecules. It is noteworthy that the success of scaling the Mulliken charges, and corresponding potential, for the same wave functions, was markedly less successful.[53] It has also been observed that the electrostatic potential calculated from a distributed multipole model is much less basis set dependent than that calculated from Mulliken charges of the same wave functions.[54] Thus, the accuracy of the scaling approximation will be very dependent on the ability of the set of atomic charges or multipoles to represent the molecular charge distribution (and therefore the electrostatic potential) accurately. Our results suggest that the common practice of scaling electrostatic interactions could have a good physical basis, although it is usually used as a pragmatic "fudge factor". This conclusion is naturally limited to the "in vacuo" electrostatic properties considered here.

It is noteworthy that the magnitude of the total dipole moment was found to be so successful as a scaling factor for the electrostatic potential. This has the advantage over an optimized scaling parameter that it will also be valid at very long range, and is independent of the choice of grid points. The total dipole moment of a large molecule is a very crude measure of its charge distribution, and totally inappropriate for use in predicting the electrostatic properties of the molecule except at very long range, so its success is somewhat unexpected. However, the success of the scaling is dependent on the direction of the dipole moment being almost independent of basis set and on the introduction of correlation making only a small difference to this direction. Scaling between potentials derived from different ab initio (or semiempirical) methods is unlikely to be so successful if this is not the case. Since the total dipole moment, and direction, for large unsymmetrical molecules is rarely reported, we have little idea whether the near independence of the dipole direction with reasonable basis set is a general phenomena, though it could be a natural consequence of a well-balanced basis set.

Although scaling may be a valid approximation in molecular modelling, it does require the use of the accurate dipole moment, which will often not be available from experiment. The large basis set MP2 calculations reported here are very expensive, and, even with future advances in computer hardware, such calculations are unlikely to become routine for significantly larger molecules in the near future. The way forward may well lie in the use of transferable electrostatic models, possibly with the inclusion of polarization effects. This will require quite large model molecules to ensure that the short-range inductive effects are the same in model and macromolecule.[35] For example, the blocked peptide used here is a suitable building block for constructing polypeptides. Fortunately, the errors in the electrostatic properties arising from using transferable models for peptides[48] can be smaller than those arising from the neglect of electron correlation. Thus, the prospects for "chemically accurate" electrostatic models for biomolecules are good, although it will require careful choice of wave function, representation, and transferability assumptions.

The outlook for realistic computer simulation is generally good, provided care is taken in interpreting the results in the light of the implicit assumptions. Although this work has highlighted an important problem in predicting the electrostatic interactions quantitatively, it has also shown that high accuracy may be achieved by scaling cheaper calculations. Nevertheless, much molecular modelling is essentially qualitative, or concerned with structures, and such properties appear to be given reliably by SCF wave functions with modest split-valence basis sets.

Acknowledgment. We acknowledge the financial support of the SERC for computer equipment under Grant GR/G11668, for computer time on the IBM 3090 at Rutherford, and for a studentship for J.S.A.

Appendix

In this discussion $\{i, j\}$ refer to occupied molecular orbitals (MOs), $\{a, b\}$ to virtual MOs, and $\{\mu, \nu\}$ to atomic orbitals.

Recently Les and Adamowicz have calculated the correlated molecular electrostatic potentials (MEPs) for some molecules at second-order Møller–Plesset theory.[33] In their work the conventional definition of the MEP as

$$V(P) = \sum_A \frac{Z_A}{R_{AP}} - \int \frac{\rho(i)}{r_{iP}} d\tau_i \qquad \text{(A1)}$$

is used and the density matrix is expanded up to second order in the correlation corrections. Assuming a closed-shell, RHF wave function, this correction can be written in terms of the density matrix as[55,56]

$$D_{\mu\nu} = D^{(0)}_{\mu\nu} + D^{(2)}_{\mu\nu} \qquad \text{(A2)}$$

where

$$D^{(0)}_{\mu\nu} = \sum_j 2C_{\mu j}C_{\nu j} \qquad \text{(A3)}$$

and $D^{(2)}_{ij}$ is defined in the molecular orbital basis through

$$D^{(2)}_{ij} = -V_{ij} \qquad D^{(2)}_{bc} = V_{bc} \qquad D^{(2)}_{ai} = -t_{ai} \qquad \text{(A4)}$$

where

$$V_{ij} = \sum_k \sum_{ab} 2(ka|ib)t^{ab}_{kj}/D^{ab}_{ik}$$
$$V_{bc} = \sum_{ij} \sum_a 2(ia|jb)t^{ac}_{ij}/D^{ab}_{ij} \qquad \text{(A5)}$$

with

$$t^{ab}_{ij} = [2(ia|jb) - (ib|ja)]/D^{ab}_{ij} \qquad D^{ab}_{ij} = \epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j \qquad \text{(A6)}$$

and the matrix t is defined through

(50) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
(51) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.
(52) Orozco, M.; Luque, F. J. *J. Comput. Chem.* **1990**, *11*, 909.
(53) Burt, C.; Reynolds, C. A.; Richards, W. G. *J. Comput. Chem.*, submitted.
(54) Sokalski, W. A.; Sneddon, S. F. *J. Mol. Graphics* **1991**, *9*, 74.
(55) Pople, J. A.; Seeger, R. *J. Chem. Phys.* **1975**, *62*, 4566.
(56) Amos, R. D. *Chem. Phys. Lett.* **1980**, *73*, 602.

$$t_{ai} = \frac{-4}{(\epsilon_a - \epsilon_i)} \sum_{jb} [\sum_c t_{ij}^{cb}(ac|jb) - \sum_k t_{kj}^{ab}(ki|jb)] \quad \text{(A7)}$$

In the present work the correlation effects at MP2 are introduced into the MEP by using the "relaxed" density matrix.[57-59] This formalism is also used by Wiberg et al.[34] This approach is compared to that of Les and Adamowicz[33] below.

The MP2 energy expression is

$$E_2 = -\sum_{ijab} (ia|jb) t_{ij}^{ab} \quad \text{(A8)}$$

The expression for the MP2 derivative was first derived by Pople et al.[60] and was later given a more efficient formalism by Handy and Schaefer.[57-61] Under a general perturbation $\lambda \rightarrow \lambda_0 + \delta\lambda$, the derivative of the total energy is given by

$$E^{(\lambda)} = \sum_{\mu\nu} [Y_{\mu\nu}^{(\mathrm{scf})} + Y_{\mu\nu}^{(2)}] h_{\mu\nu}^{\lambda} + \sum_{\mu\nu} [W_{\mu\nu}^{(\mathrm{scf})} + W_{\mu\nu}^{(2)}] S_{\mu\nu}^{\lambda} + \sum_{\mu\nu\lambda\sigma} [\Gamma_{\mu\nu\lambda\sigma}^{(\mathrm{scf})} + \Gamma_{\mu\nu\lambda\sigma}^{(2)}] (\mu\nu/\lambda\sigma)^{\lambda} \quad \text{(A9)}$$

The matrices **Y**, **W**, and **Γ** are independent of the nature of the perturbation and depend solely on the form of the wave function (e.g., expressions involving integrals and orbital energies). Only **Y** is required for one-electron properties. $\mathbf{Y}^{(\mathrm{scf})}$ is equal to $\mathbf{D}^{(0)}$ while $\mathbf{Y}^{(2)}$ is defined in the MO basis through

$$Y_{ij}^{(2)} = -V_{ij} \qquad Y_{bc}^{(2)} = V_{bc} \qquad Y_{ai}^{(2)} = -Z_{ai} \quad \text{(A10)}$$

where **Z**, the response vector, is calculated through

$$Z_{ai}(\epsilon_a - \epsilon_i) + \sum_{bj} A_{aibj} Z_{bj} = L_{ai} \quad \text{(A11)}$$

and where

$$L_{ai} = -4 \sum_{jb} [\sum_c t_{ij}^{cb}(ac|jb) - \sum_k t_{kj}^{ab}(ki|jb)] + \sum_{bc} V_{bc} A_{bcai} - \sum_{jk} V_{jk} A_{jkai} \quad \text{(A12)}$$

and

$$A_{pqrs} = 4(pq|rs) - (pr|qs) - (ps|qr) \quad \text{(A13)}$$

Hence by comparing (A4) and (A10) it is seen that, at second-order in the Møller–Plesset series, the two definitions of the MEP differ only by the occupied-virtual block of the second-order density matrix, t_{ai} or Z_{ai}. Further, if the contractions with the **A** matrix in eq A11 and A12 are not computed, then it follows that these matrix elements become identical. Thus it follows[59] that the derivative formalism will give the expectation result if no iteration is performed at the coupled Hartree–Fock stage in the calculation of **Z**.

Various authors have argued that the properties of a molecule are more accurately defined as energy derivatives than as expectation values, e.g., refs 62 and 63. In the case of the MEP, we note that the potential at a point should be the path integral of force acting on a unit test charge brought from infinity to that point. The relaxed density matrix is defined such that the forces are correct; this is not true of the expectation value (Hellmann–Feynman) definition. Thus the energy derivative definition of the MEP should be more accurate.

The two methods were numerically compared by performing calculations on *N*-acetylalanine *N'*-methylamide with the DZ basis.[39] The expectation formalism gave a dipole moment direction very close to the derivative formalism, the scalar product between the two dipoles corresponding to an angle of 0.94°, but with a smaller dipole magnitude (2.76 ea_0 compared to 2.88 ea_0). The positions of the maxima and minima of the electrostatic potential, on a surface 1.4 Å from the van der Waals surface, were the same for both methods. The rms value of

$$\Delta V = V(\text{expectation,DZ,MP2}) - V(\text{derivative,DZ,MP2})$$

is only 2.7 kJ mol^{-1} with a maximum absolute error of 7.9 kJ mol^{-1}. Thus the difference between the two methods is small but not negligible. Wiberg et al.[34] also find small, but occasionally significant differences between the two approaches.

Registry No. AcNHCH(CH_3)CONHMe, 22715-68-0.

(57) Rice, J. E.; Amos, R. D. *Chem. Phys. Lett.* **1985**, *122*, 585.
(58) Simandiras, E. D.; Amos, R. D.; Handy, N. C. *Chem. Phys.* **1987**, *114*, 9.
(59) Trucks, G. W.; Salter, E. A.; Sosa, C.; Bartlett, R. J. *Chem. Phys. Lett.* **1988**, *147*, 359.
(60) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem.* **1979**, *S13*, 225.
(61) Handy, N. C.; Schaefer, H. F. *J. Chem. Phys.* **1984**, *81*, 5031.
(62) Nerbrant, P-O; Roos, B.; Sadlej, A. J. *Int. J. Quantum Chem.* **1979**, *15*, 135.
(63) Sadlej, A. J. *J. Chem. Phys.* **1981**, *75*, 320.
(64) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97.