Molecular Electrostatic Potentials and Partial Atomic Charges from Correlated Wave Functions: Applications to the Electronic Ground and Excited States of 3-Methylindole

John D. Westbrook, Ronald M. Levy, and Karsten Krogh-Jespersen*

Department of Chemistry, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08903

Received 13 December 1991; accepted 13 April 1992

Procedures have been developed to generate molecular electrostatic potentials based on correlated wave functions from *ab initio* or semiempirical electronic structure programs. A new algorithm for point-wise sampling of the potential is described and used to obtain partial atomic charges via a linear, least squares fit between classical and quantum mechanical electrostatic potentials. The proposed sampling algorithm is efficient and promises to introduce less rotational variance in the potential derived partial charges than algorithms applied previously. Electrostatic potentials and fitted atomic charges from *ab initio* (HF/6-31G* and MP2/6-31G*) and semiempirical (INDO/S; HF, SECI, and SDCI) wave functions are presented for the electronic ground (S_0) and excited (^1L_b, ^1L_a) states of 3-methylindole.

INTRODUCTION

The concept of the electrostatic potential as a quantum mechanical observable (the expectation value of the one-electron operator $\mathbf{r}^{-1}$) probing the molecular charge distribution and its usage as a tool for describing molecular interactions and chemical reactivity is well documented.[1,2] The development of procedures to generate classical partial atomic charges from the quantum mechanical electrostatic potential provides well-defined methods for obtaining electrostatic interaction parameters required in molecular mechanics and dynamics simulations.[3-5] The quantum mechanical electrostatic potential is evaluated from the electronic wave function at a number of spatial points surrounding the molecule, and an optimal set of partial atomic charges is then obtained from a least squares fit of the classical electrostatic potential exerted by such charges to the quantum mechanical potential. The sampling of electrostatic potential points is usually made at and beyond the molecular van der Waals surface, and hence medium- to long-range (molecular dipole type) interactions are reasonably well represented by the derived atomic monopoles. To the best of our knowledge, previous applications of these atomic charge-fitting procedures have been restricted to electronic ground states with wave functions obtained at the Hartree–Fock (HF) level of approximation (*ab initio* or semiempirical).

We are modeling photophysical processes in solution through a combination of classical molecular dynamics simulation and quantum mechanical electronic structure techniques.[6-8] Electrostatic interactions are dominant for polar solutes in polar media and are in our simulations mediated in the traditional way via atom-centered partial charges. Proper description of the electrostatic interactions between any electronic state of the solute and the solvent are essential for the validity of the simulations. The molecules presently of interest to us are too large and of too low symmetry to permit meaningful *ab initio* treatments of their electronically excited states; however, it is possible to carry out *ab initio* calculations on the ground states of these species with large basis sets and some extent of electron correlation included. Wave functions for the electronically excited states are instead typically obtained from a configuration interaction calculation within a semiempirical electronic structure (INDO/S) framework.[9,10] We thus require atomic charges from wave functions obtained by methods that go beyond the HF level.

We outline here our approach to obtaining partial atomic charges from correlated wave functions, appropriate to the electronic ground or an excited state, and present results for 3-methylindole (3-MeIn). The procedures rely primarily upon the availability of the relevant one-electron reduced density matrix[11] and a good point sampling scheme. We encountered some problems with the point selection algorithms most commonly used in the past,[12,13] including substantial rotational vari-

*Author to whom all correspondence should be addressed.

CCC 0192-8651/92/080979-11$04.00

ance displayed by the potential fitted partial charges. We briefly describe a new sampling scheme that promises to rectify most of these deficiencies and be quite efficient as well.

METHODOLOGY

The quantum mechanical molecular electrostatic potential at a spatial point r may be evaluated from the formula

$$V(r) = \sum Z_A/|R_A - r| - \int \rho(r')/|r' - r| \, dr'$$
$$= \sum Z_A/|R_A - r| - \sum p_{\mu\nu} \times \int \phi_\mu(r')\phi_\nu(r')/|r' - r| \, dr' \quad (1)$$

where Z_A is the core charge of nucleus A at position R_A, $\rho(r')$ is the electron density at the point r', and $p_{\mu\nu}$ is the element of the one-electron reduced density matrix between the atomic basis functions ϕ_μ and ϕ_ν. The reduced density matrix may be computed from an independent particle (HF type) or correlated (configuration interaction, perturbation theory) wave function obtained from an *ab initio* or a semiempirical electronic structure method.[11] In the present work, *ab initio* HF and correlated (Møller–Plesset second order perturbation theory, MP2) density matrices were extracted from the checkpoint files created by the Gaussian 90 program system.[14] The semiempirical correlated density matrices (configuration interaction with single excitations, SECI, or single and double excitations, SDCI) were obtained using the projective reduction configuration interaction procedure proposed by Boys and coworkers.[15] Detailed descriptions of the implementation of this procedure are available from several sources[16a–d]; our algorithm follows closely the treatment presented by Reeves.[16a]

Two computer programs, CHELP (QCPE # 524)[17] and ESP (QCPE # 589),[18] are widely available for sampling the electrostatic potential at various points around a molecule and subsequently obtaining the atom-centered, partial charges (Q_A) by linear, least-squares fitting of the classical electrostatic potential exerted by such charges [$E(r, Q_A) = \Sigma\, Q_A/|R_A - r|$] to the quantum mechanical electrostatic potential [$V(r)$]. The primary input to these programs consists of the atomic basis sets used to evaluate the integrals appearing in (1), the elements of the one-particle reduced density matrix, and specification of the molecular geometry.

Initial attempts to use either CHELP or ESP for highly symmetrical molecules such as benzene (D_{6h}) or 9,9′-bianthryl (D_{2d}) led to severe symmetry breaking in the potential fitted charges, Q_A. The partial atomic charges exhibited lower symmetry than the quantum mechanical electrostatic potential from which they were derived; it was verified that $V(r)$ correctly displayed the full molecular symmetry with very high numerical accuracy. Furthermore, the magnitudes of the fitted charges and the extent of asymmetry were strongly dependent upon the orientation of the molecule within the global coordinate system. These orientational problems appear to be exacerbated as the molecular size increases and partial charge differences up to 0.3 e arising solely from rotational variance have been encountered for the C-methyl atom in 3-MeIn (Fig. 1) using standard sampling criteria implemented in CHELP. Rotational variance, although seemingly to a smaller extent, has been noted in the CHELP program before[19,20]; the effect is also present, but less pronounced, in the ESP code. A characteristic of both programs is that the sampling of points is carried out on spheres in local, atom-centered coordinate systems that do not alter their orientation if the molecule is rotated as a whole. The approach to the rotational variance problem most often taken in previous studies has been to alter the sampling procedure and at the same time dramatically increase the number of sampling points relative to previously recommended levels. The defaults set in either CHELP or ESP are coarse and typically lead to a total molecular sampling space of only a few hundred points. Breneman and Wiberg use approximately 5000–6000 points for a molecule as small as formamide[19] and Woods et al.[20] recommend 1500–2500 points per atom (!). A judicious increase in the number of points will certainly diminish the rotational variance; however, the rate of improvement is slow with the selection algorithms used in both CHELP and ESP from even a substantial increase (by a factor of 5–10) in the total number of sampling points. Reverting to the original sampling scheme of Cox

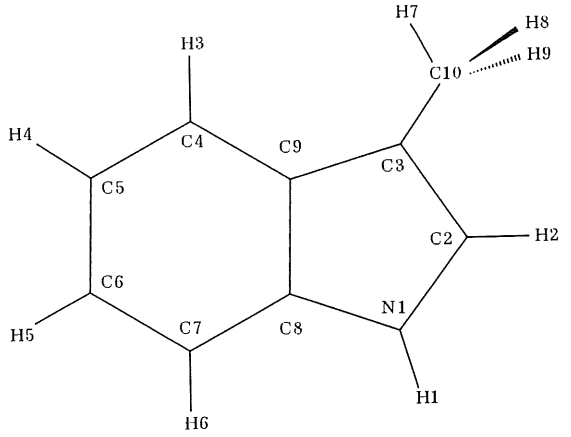


Figure 1. Atomic numbering scheme for 3-methylindole.

and Williams[4] based on a grid in a cubic box has also been recommended.[19] We find that with this scheme, which does appear to be less rotationally variant than the atom sphere based schemes of CHELP and ESP, the charges on benzene developed from more than 40,000 electrostatic potential points in an AM1 calculation using the MOPAC program[18,21] still show a "4-2" pattern of broken symmetry (D_{2h}); 1000 points and a "poor" choice of global molecular orientation may lead to different charges on all the atoms in benzene. For large molecules, the computer time needed to evaluate perhaps 50,000 electrostatic potential points is considerable, even with the minimal basis sets used in semiempirical methods, and much longer than the time required for the basic electronic structure calculation itself. We have therefore developed a sampling scheme that promises to significantly reduce the rotational variance relative to the schemes currently in use without a concomitant significant increase in computational cost.

Points for the electrostatic potential fitting procedures are collected using the following prescription. First, a number of points are chosen on a unit sphere at intersecting lines of longitude and latitude. Using lines of longitude and latitude separated by 15° in azimuthal and polar angles, respectively, yields 266 points of intersection. Our sampling points are distributed in a geometrically uniform manner, since the same number of points is placed on any circumference. The sampling procedure in ESP assigns a constant number of points per unit distance, thus emphasizing the equatorial sphere region, whereas the sampling procedure in CHELP simply is too coarse, generating only 14 points per sphere. The unit spheres are then scaled by the van der Waals radius of each atomic center and by an additional surface scale factor. The resulting spheres are positioned at the atomic centers with no rotation of the local sphere coordinates relative to the global coordinate system. When all the spheres are in place, the interior points (points that lie inside a sphere belonging to another atom) are discarded. In the work reported here we use four surfaces, obtained by scaling the atomic van der Waals radii by factors of 1.2, 1.4, 1.6, and 1.8, respectively. The resulting surfaces surrounding 3-MeIn contain 1923, 1662, 1467, and 1284 points, respectively, a total of about 6000 points. The electrostatic potential is evaluated at these points and the partial atomic charges are then readily obtained using a Lagrange multiplier based linear, least-squares fit with the total molecular charge and (optionally) the nonzero components of the dipole moment as constraints.[12,22] Thus, the function

$$Y(Q_A) = \sum [V(r) - E(r, Q_A)]^2 \qquad (2)$$

expressing the squares deviation between the quantum mechanical and classical electrostatic potentials over the sampled points, is minimized subject to the constraint that $\Sigma\, Q_A = Q_T$ = total molecular charge.

Our sampling procedure makes it straightforward to evaluate the rotational variance due to the orientation of the sampling points within the global coordinate system. The procedure is invariant to any rotation around the principal coordinate axes that can be expressed as an integral multiple of the angular separation used in selecting points on the unit sphere, 15° in the present case. Maximum rotational variance typically occurs from a rotation of the molecular coordinate system by one half of this angular separation; such a coordinate rotation (i.e., 7.5°) for 3-MeIn results in changes of at most 0.02 units in the calculated partial atomic charges. It is important to point out, however, that although the derived partial atomic charges reproduce the quantum mechanical electrostatic moments with great accuracy (even without the dipole moment constraint included in the fitting procedures), the classical electrostatic potential itself may deviate from the quantum mechanical potential by as much as 25% for points on the innermost sampling surface. We have attempted to reduce the error by including additional constraints on higher electrostatic moments (quadrupole, etc.) or by seeking to diminish the errors regionally in certain parts of the molecule using nonlinear weighted least-squares fitting procedures. Although this additional effort did not significantly improve our results with 3-MeIn, we have observed that the ability to optimize fitting procedures to specific regions of a molecule may be particularly useful when applied to molecules containing very polar functional groups.

When evaluating the integrals and the density matrix elements appearing in (1), the exact representation of the atomic basis set (gaussians) is obvious when an *ab initio* method is used. The actual analytical representation of the atomic basis set plays only a small role in the semiempirical, all valence electron methods (MNDO, AM1, INDO/S, etc.) but is usually taken to be of the Slater-type; however, the SCF equations are solved using the zero differential overlap (ZDO) approximation with a unit overlap matrix. We follow the recommendations made by other researchers for evaluating properties from semiempirical wave functions[23-26] and compute the one-electron reduced density matrix over molecular orbitals deorthogonalized[27] with an overlap matrix based on the STO-6G expansions of Slater orbitals given by Hehre et al.[28] The electrostatic potential integrals in (1) are also evaluated using the STO-6G basis set.

We have interfaced the electrostatic potential/sampling schemes with standard *ab initio* codes

(Gaussian 90, GAMESS)[14,29] as well as the semiempirical MOPAC[21] and Excited State Properties Package (ESPPAC)[30] programs. ESPPAC is an electronic structure program developed at Rutgers University based on the INDO model Hamiltonian with emphasis on the computation of properties for electronically excited states.

RESULTS AND DISCUSSION

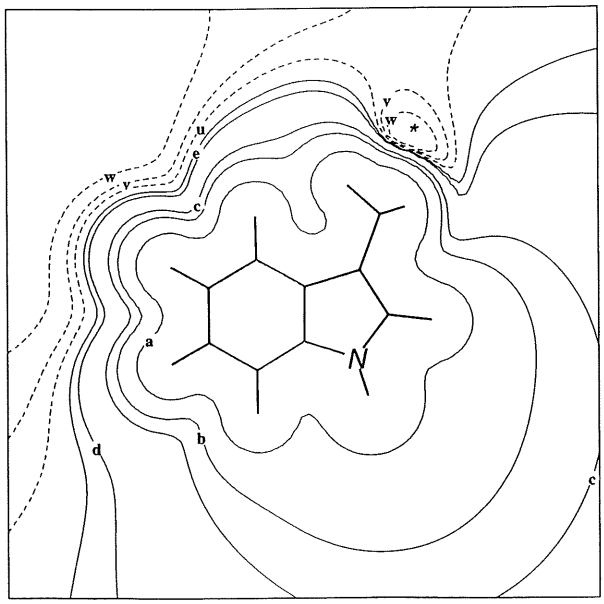
Electronic Ground State of 3-MeIn

The geometry of 3-MeIn was fully optimized at the HF level with the 6-31G* basis set and Gaussian 90 (HF/6-31G*//6-31G*).[14] The electrostatic potential was sampled as outlined above using reduced one-electron density matrices from both uncorrelated (HF) and correlated (MP2) wave functions. A plot of the potential in the molecular plane obtained with the MP2/6-31G* electron density is shown in Figure 2(a). The in-plane potential near the molecule is dominated by the hydrogen atoms and solid contours of positive electrostatic potential (repulsive to a positive test charge) envelop the entire molecule.[1] Only at distances of ~2.0–2.5 Å from the core of any atom and beyond does the potential become attractive in the parts of the molecular plane that contain the negative end of the molecular dipole. The computed magnitude of the permanent ground state dipole moment is 2.08 D and the orientation is roughly parallel to a line passing through atoms C5 and C7 with the positive end toward C7. The experimental value[31] for the dipole moment of indole is 2.13 D and the dipole moment of 3-MeIn should be very similar. A local minimum in the electrostatic potential (~ −1.0 kcal/mol) is located in the vicinity of the C-methyl atom.*

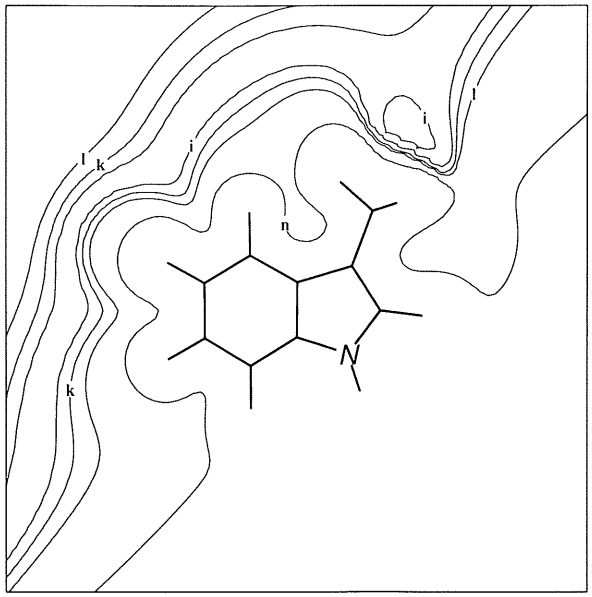
A plot of the in-plane electrostatic potential derived from the HF/6-31G* density is very similar to the plot presented as Figure 2(a) and differences are hardly noticed even when the two plots are placed next to each other. A simple difference plot [$\Delta V(r) = V_{MP2}(r) - V_{HF}(r)$] appears very noisy due to numerous sign changes occurring near the boundary between positive and negative values of the electrostatic potential. We show instead as Figure 2(b) an absolute (percentage) difference plot defined by

$$\Delta V(r) = \mathrm{abs}\{[V_{MP2}(r) - V_{HF}(r)]/ 0.5*[V_{MP2}(r) + V_{HF}(r)]\} \quad (3)$$

It can be seen from comparisons of Figures 2(a) and 2(b) that the largest percentage differences between the MP2 and HF based electrostatic potentials [contour *i* in Fig. 2(b)] lie along the boundary between positive and negative potential contours. The electrostatic potential in the molecular plane is more attractive (to a positive test charge) at the MP2 than at the HF level and the boundary is uni-



(a)



(b)

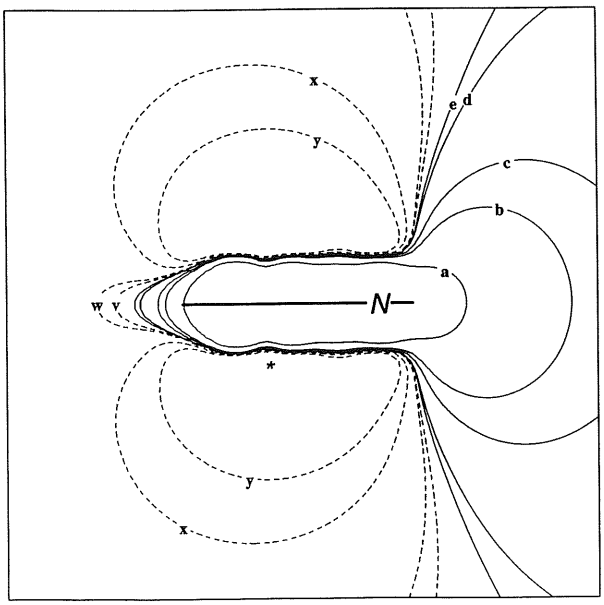
Figure 2. (a) Molecular electrostatic potential for 3-methylindole computed from a MP2/6-31G*//HF/6-31G* wave function and plotted in the molecular plane. The contour levels (in Hartrees; 1 Hartree = 627.5 kcal/mol) are as follows: **a** = 0.05; **b** = 0.01; **c** = 0.005; **d** = 0.001; **e** = 0.0005; and **u** = −0.0001; **v** = −0.0005; and **w** = −0.001. The location of the local minimum is indicated by "*." (b) Electrostatic potential absolute difference, $\Delta V(r)$, defined according to eq. (3) and plotted in the molecular plane. The contour levels (%) are as follows: **i** = 150; **k** = 80; **l** = 50; and **n** = 10.

*An anonymous referee has kindly pointed out that the presence of negative potentials in the cones of methyl groups was discovered by Politzer and Daiker[32] and discussed in more detail by them in 1981.[33]

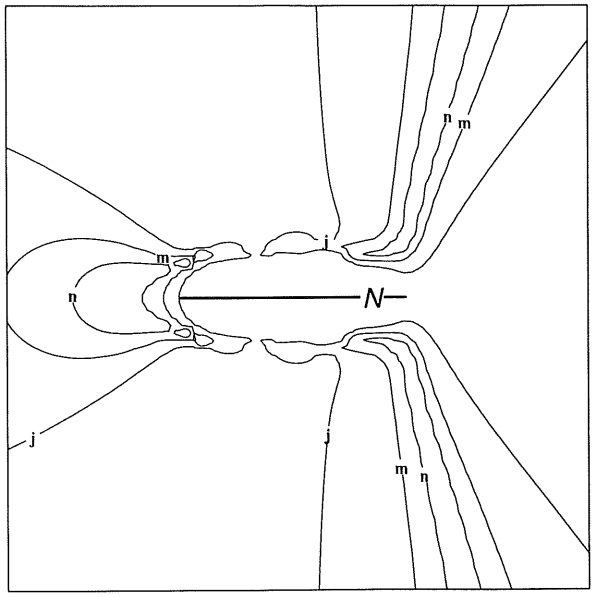
formly 0.2–0.5 Å closer to the molecule at the correlated level.

In Figures 3(a) and 3(b), we present corresponding plots of the MP2/6-31G* based electrostatic potential and the absolute difference, $\Delta V(r)$, in a plane perpendicular to the molecular plane aligned so that the intersection of the two planes contains the dipole moment vector as well as the center of mass. With the chosen orientation the perpendicular plane roughly bisects the C4—C9, C8—C9, and C8—N1 bonds. Again, the entire molecule is enveloped by solid positive potential contours that change rapidly in value with increasing distance from the molecular plane [Fig. 3(a)]. The potential turns negative early (~1.3 Å) below or above the six-membered ring but the space below (above) and outside the nitrogen atom is consistently characterized by positive electrostatic potential contours. As expected, the attractiveness of the potential is much more pronounced perpendicular to the molecular plane than in the molecular plane, and the minimum value in the chosen perpendicular plane (~ −23 kcal/mol) is reached at a point located approximately 1.7 Å below (or above) the molecular plane, slightly inside the six-membered ring near the midpoint of the C4—C9 bond. The largest relative differences between the MP2 and HF electrostatic potentials occur again along the boundary between the positive and negative contours [contour **j** in Fig. 3(b)]. Correlation also renders the electrostatic potential perpendicular to the molecular plane more attractive, in regions both directly above (below) the molecular plane and well outside the molecular boundary.

Sets of classical point charges were fit to the HF and to the MP2 electrostatic potentials (Table I). The differences between the two charge sets are small for this molecule, amounting to at most 0.05 e (C2, C4, C5, N1), which is fully consistent with the small differences observed in the electrostatic potentials or in the directions and magnitudes of the dipole moments [1.94 D (HF); 2.08 D (MP2)]. The charges obtained from a standard Mulliken population analysis[34] (Table I) tend to be numerically larger than the electrostatic potential fitted charges and so do the changes from the HF to the MP2 level, but the directions of change between the two charge sets are mostly the same. However, the classical molecular dipole moments for 3-MeIn computed from the Mulliken atomic charges and the optimized molecular geometry do not reproduce the properties of the quantum mechanical dipole moments well.[4] Their magnitudes (HF: 0.82 D; MP2: 0.95 D) are only about one half as large as the quantum mechanical dipole moments computed from the appropriate wave functions and, furthermore, these classical dipole moments are rotated nearly 90° counterclockwise



(a)



(b)

Figure 3. (a) Molecular electrostatic potential for 3-methylindole computed from a MP2/6-31G*//HF/6-31G* wave function and plotted in a plane perpendicular to the molecular plane. The contour levels (in Hartrees; 1 Hartree = 627.5 kcal/mol) are as follows: **a** = 0.05; **b** = 0.01; **c** = 0.005; **d** = 0.001; **e** = 0.0005; and **y** = −0.01; **x** = −0.005; **w** = −0.001; **v** = −0.0005. The location of the minimum is indicated by "*." For clarity, the projections of two methyl C—H bonds have been omitted from the edge-on view of the molecule. (b) Electrostatic potential absolute difference, $\Delta V(r)$, defined according to eq. (3) and plotted in the molecular plane. The contour levels (%) are as follows: **j** = 90; **m** = 30; and **n** = 10. For clarity, the projections of two methyl C—H bonds have been omitted from the edge-on view of the molecule.

Table I. Partial atomic charges for the ground state of 3-methylindole computed at the *ab initio* HF/6-31G* and MP2/6-31G* levels and with the semiempirical MNDO model.[a]

Atom[b]	$Q_{V,HF}$[c]	$Q_{V,MP2}$[c]	$Q_{M,HF}$[d]	$Q_{M,MP2}$[d]	Q_{MNDO}[c,e]
N1	−0.497	−0.448	−0.839	−0.702	−0.547
C2	−0.134	−0.135	0.036	−0.014	−0.129
C3	−0.065	−0.030	−0.018	0.040	−0.116
C4	−0.188	−0.192	−0.198	−0.199	−0.076
C5	−0.187	−0.132	−0.226	−0.182	−0.229
C6	−0.065	−0.086	−0.210	−0.196	−0.043
C7	−0.336	−0.278	−0.218	−0.181	−0.335
C8	0.298	0.249	0.333	0.264	0.229
C9	0.018	0.018	−0.067	−0.015	−0.117
C10	−0.052	−0.114	−0.500	−0.543	−0.011
H1	0.378	0.370	0.389	0.365	0.475
H2	0.179	0.159	0.211	0.179	0.238
H3	0.153	0.133	0.201	0.166	0.148
H4	0.121	0.102	0.191	0.162	0.134
H5	0.114	0.101	0.194	0.162	0.124
H6	0.165	0.143	0.197	0.165	0.185
H7	0.040	0.053	0.175	0.177	0.036
H8	0.029	0.044	0.174	0.176	0.017
H9	0.029[f]	0.044[f]	0.174	0.176	0.017
μ[g]	1.941	2.078	1.941	2.078	1.910
μ[h]	1.941[i]	2.078[i]	0.819	0.946	1.910[i]

[a] HF/6-31G*//6-31G* geometry used throughout.
[b] The atomic numbering scheme is shown in Figure 1.
[c] Charges fitted to electrostatic potential.
[d] Charges from Mulliken population analysis.
[e] Charges scaled by a multiplicative factor of 1.42.[18]
[f] Equivalent by symmetry to H8 (C_s point group) but treated independently in the least-squares fit. The fitted charges on H8 and H9 are in fact identical to at least 6 decimals.
[g] Quantum mechanical dipole moment (Debye) computed from the wave function.
[h] Classical dipole moment (Debye) computed from the atomic charges and the optimized HF/6-31G* geometry.
[i] The quantum mechanical dipole moment magnitudes were used as constraints in fitting the partial charges.

relative to the quantum mechanical dipole moments. Comparison to available experimental data[31] (see above) shows that the computed quantum mechanical dipole moments for 3-MeIn are quantitatively correct. Thus, direct application of the Mulliken charges in molecular simulations would produce grossly incorrect electrostatic interactions. Also included in Table I are the potential derived charges computed from the semiempirical MNDO wave function for 3-MeIn according to the sampling and fitting procedures specified in the methodology section, scaled up by the factor of 1.42 found in a previous study to lead to good agreement between MNDO and HF/6-31G* potential derived partial atomic charges on a set of small test molecules.[13,18] The overall agreement between the scaled MNDO and the HF/6-31G* charges for this large, extended molecule appears satisfactory, although detailed comparisons of columns 1 and 5 in Table I reveal that several substantial differences (up to ~0.1 e) do exist. The optimal scale factor in 3-MeIn extracted from a least-squares fit between MNDO and HF/6-31G* charges derived with our sampling procedures is actually 1.36.

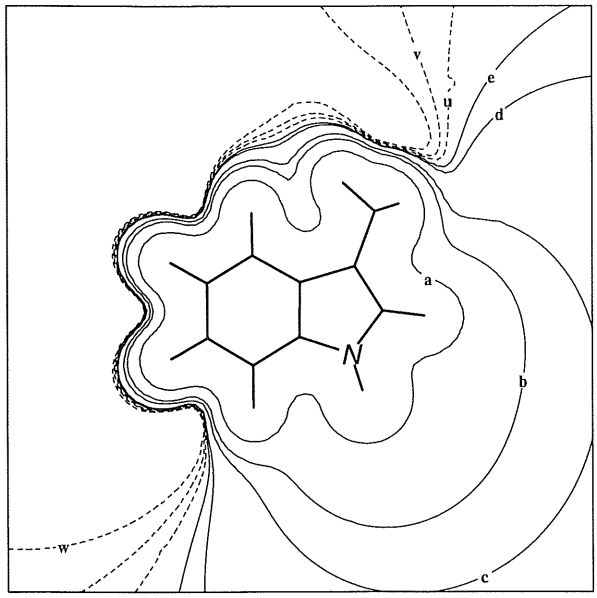
Electronically Excited States of 3-MeIn

Indole is the absorbing chromophore of the amino acid tryptophan and is the dominant naturally occurring fluorophore in proteins.[35] The photophysics of indoles have been under scrutiny for years; a detailed theoretical description of several excited state properties computed by semiempirical methods may be found in the recent work by Callis.[36] Basic spectroscopic features include the presence of two close-lying electronically excited singlet states in the near UV region, conventionally denoted ^1L_a and ^1L_b in an extension of Platt's nomenclature.[37] The indole ^1L_a state has a much larger dipole moment (estimated ~5 D)[38] than the ground state (2.1 D)[31] or the ^1L_b state (~2.3 D).[39] In the gas phase under supersonic jet conditions, the ^1L_b state is slightly lower in energy than the ^1L_a state.[40] However, due to the significant differences in polarity of the two excited states, the state ordering is readily reversed in polar solvent and the actual ^1L_a-^1L_b energy separation and many other excited state features show strong dependence on the solvent polarity.[41] The photophysical properties of indoles thus reflect both the nature of the chromophore and the surrounding medium.

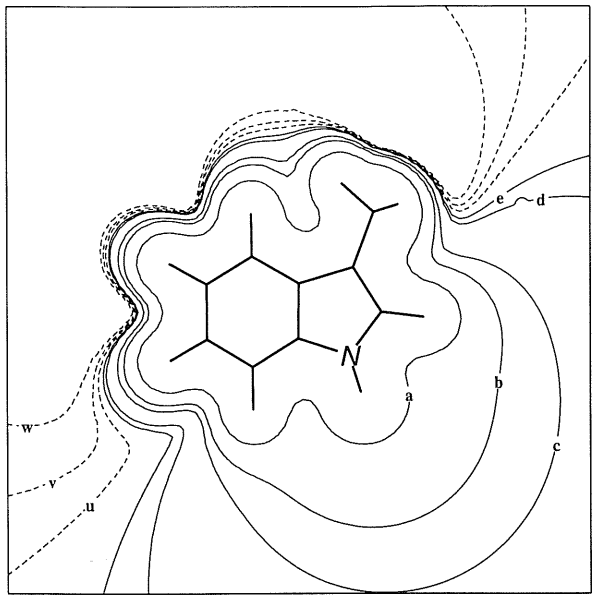
INDO/S calculations with configuration interaction including single and double excitations relative to the ground state reference determinant were carried out on 3-MeIn at the optimized HF/6-31G* geometry; the semiempirical parameters used were those recommended in ref. 9. The excitation space consisted of all possible single excitations (SECI), or all possible single excitations plus the double excitations (SDCI) derived from the 10 or 20 molecular orbitals bracketing the Fermi level [5(10) "up," 5(10) "down"]. The one-electron reduced density matrices for the ground, ^1L_b, and ^1L_a states were assembled in the deorthogonalized basis and used to compute the electrostatic potentials (cf. methodology section). Constant potential contours in the plane of the molecule from the largest calculation (all single and 10 "up," 10 "down" double excitations for a total of 4115 configurations) are shown as Figures 4(a) (S_0), 4(b) (^1L_b), and 4(c) (^1L_a); the contour values used for Figure 2 were also applied in Figures 4(a)–4(c) to facilitate comparisons.

Comparison of Figures 4(a) and 2(a) shows that the electrostatic potential for the ground state of 3-MeIn created from the semiempirical INDO/S SDCI and the *ab initio* MP2/6-31G* wave functions possess the same gross qualitative features but also that there are some quantitative differences. The

boundary between positive and negative values of the electrostatic potential is reached considerably closer (~0.5 Å) to the molecule in the semiempirical calculation [Fig. 4(a)] and the local minimum near the methyl group is absent. The positive contours of Figures 4(a) and 2(a) are nearly superimposable, in particular for the innermost, high energy contour but the INDO/S electrostatic potential diminishes rapidly with distance and changes quite abruptly near the boundary region. The MP2/6-31G* electrostatic potential changes gradually across this region whereas that of the INDO/S calculation has the appearance of a "wall." The 3-MeIn dipole moment computed in the INDO/S SDCI calculation is 1.74 D. It is slightly smaller than the dipole moment from the MP2/6-31G* calculation (2.08 D), which inherently would tend to make the electrostatic potential more slowly varying in the INDO/S calculation. It is thus more likely that the potentials reflect the substantial differences in the natures of the atomic basis sets employed in the two types of calculations. The spatially limited and inflexible minimal basis set (STO-6G) in the semiempirical method is inferior to the *ab initio* basis set of split valence plus polarization quality (6-31G*). This manifests itself in comparisons of electrostatic potentials computed with HF wave functions[26,42] and, as the present results demonstrate, also at the correlated levels. The diffuseness of the INDO/S atomic 2*p*-orbitals could be readily increased and perhaps render the electrostatic potential maps quantitatively in agreement at long distances, but this might upset the delicate balance among the semiempirical parameters and influence the computation of other molecular properties.

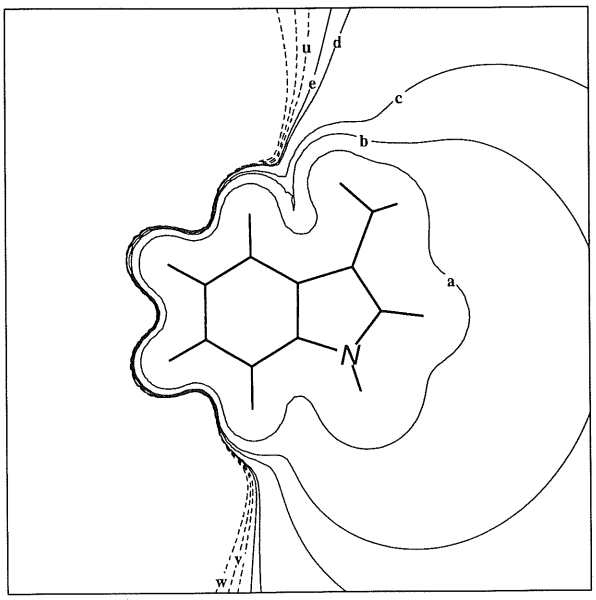


(a)



(b)

Figure 4. Molecular electrostatic potentials for the S_0 (a), ^1L_b (b), and ^1L_a (c) states of 3-methylindole computed from INDO/S SDCI wave functions and plotted in the molecular plane. The contour levels (in Hartrees; 1 Hartree = 627.5 kcal/mol) are as follows: **a** = 0.05; **b** = 0.01; **c** = 0.005; **d** = 0.001; **e** = 0.0005; and **w** = −0.001; **v** = −0.0005; **u** = −0.0001.



(c)

Figure 4. (continued)

The electrostatic potentials of the ^1L_b and S_0 states are overall similar in appearance [Figs. 4(b) and 4(a)] but the dipole moment is slightly larger in magnitude and rotated (about 10° counterclockwise) in ^1L_b relative to S_0.[36,39] The computed change $[\mu(^1L_b) - \mu(S_0)]$ for 3-MeIn is 0.38 D (Table II) which may be compared to an absolute change determined by Stark effect measurements of 0.14 D in the parent indole molecule.[39] The substantial

Table II. Partial atomic charges computed by electrostatic potential fitting of INDO/S wave functions for the S_0, ^1L_b, and ^1L_a states of 3-methylindole.[a]

Atom[b]	S_0^{SDCI}	$^1L_b^{SDCI}$	$^1L_a^{SDCI}$	S_0^{HF}	$^1L_b^{SECI}$	$^1L_a^{SECI}$
N1	−0.392	−0.385	−0.318	−0.386	−0.338	−0.357
C2	−0.073	−0.042	−0.002	−0.080	−0.044	0.000
C3	−0.009	0.035	0.250	−0.006	0.039	0.177
C4	−0.131	−0.133	−0.186	−0.134	−0.196	−0.159
C5	−0.046	−0.018	−0.039	−0.058	−0.005	−0.096
C6	−0.019	−0.013	−0.054	−0.020	−0.047	−0.031
C7	−0.166	−0.172	−0.315	−0.179	−0.213	−0.250
C8	0.198	0.173	0.289	0.200	0.204	0.260
C9	0.065	0.048	−0.133	0.069	0.016	−0.079
C10	−0.028	−0.057	−0.116	−0.026	−0.046	−0.080
H1	0.293	0.291	0.282	0.294	0.288	0.286
H2	0.094	0.087	0.070	0.100	0.089	0.074
H3	0.064	0.052	0.054	0.069	0.063	0.058
H4	0.027	0.014	0.021	0.032	0.020	0.037
H5	0.028	0.018	0.017	0.032	0.026	0.026
H6	0.068	0.059	0.083	0.073	0.066	0.075
H7	0.009	0.013	0.030	0.008	0.010	0.017
H8	0.009	0.015	0.034	0.006	0.011	0.022
H9[c]	0.009	0.015	0.034	0.006	0.011	0.022
μ[d]	1.747	2.133	5.394	1.838	2.939	4.319
μ[e]	1.215	1.672	4.922	1.294	2.508	3.977

[a]HF/6-31G*//6-31G* geometry used throughout.
[b]The atomic numbering scheme is shown in Figure 1.
[c]Equivalent by symmetry to H8 (C_s point group) but treated independently in the least-squares fit. The fitted charges on H8 and H9 are in fact identical to at least 6 decimals.
[d]Quantum mechanical dipole moment (Debye) computed from the wave function.
[e]Classical dipole moment (Debye) computed from the INDO/S ZDO atomic charges and the optimized HF/6-31G* geometry.

increase in dipole moment magnitude (~3.5 D) and reorientation (~25° counterclockwise) in the ^1L_a state relative to S_0 is readily evident in Figure 4(c). The computed dipole moment is 5.39 D for ^1L_a (1.74 D for S_0) and the potential around the N atom is far more repulsive (to a positive test charge) in the ^1L_a state [Fig. 4(c)] than in S_0 [Fig. 4(a)]; also, the attractive part of the potential has been pulled in closer to the atoms in the six-membered ring.

In Table II, we present the electrostatic potential derived partial atomic charges for the three states from the INDO/S calculations. We present an internally consistent set of SDCI charges for the S_0, ^1L_b, and ^1L_a states (excitation space as described above), but also, for comparison purposes and to assess the effects of the double excitations, the HF charges for S_0 and the SECI charges for ^1L_b and ^1L_a.* The largest differences between S_0^{SDCI} and S_0^{HF} fitted charges are only 0.01 e (C5 and C7) so correlation within the INDO/S framework has not changed the ground state wave function of 3-MeIn appreciably, at least as measured by the electrostatic potential. The partial atomic charges from INDO/S are without exception smaller in absolute value than those presented in Table I (*ab initio* or scaled MNDO) even though all the computed dipole moments fall within a very narrow range [e.g., at the HF level: 1.84 D (INDO/S), 1.94 D (*ab initio*), 1.91 D (MNDO)]. The partial atomic charges in the S_0 and ^1L_b states are similar as anticipated from the large uniformity in electrostatic potential maps [Figs. 4(a) and 4(b)], dipole moment magnitudes (Table II), and orientations. Larger changes in the atomic charges are observed upon excitation to the ^1L_a state. In particular, the electron densities decrease on atoms N1, C3, and C8, and they increase on C4, C7, and C9. A net transfer of charge takes place from the five- to the six-membered ring in the ^1L_a state, which accounts for the increase and reorientation of the dipole moment relative to the ground state.[36] At the SECI level, the polarity of the ^1L_b state is overestimated (2.94 D) and that of ^1L_a seems underestimated (4.32 D).† Only the SDCI calculations produce dipole moments for ^1L_b slightly larger than the ground state values as demanded by the available experimental data.[39]

The potential derived classical charges pre-

*The results from the INDO/S SDCI calculations with 5 "up," 5 "down" double excitations are indistinguishable from those calculated with 10 "up," 10 "down" double excitations in terms of both electrostatic potential and fitted atomic charges.

†The charges used in ref. 8 were obtained at the HF or SECI levels with the standard ESP sampling scheme recommended in refs. 13 and 18 on an HF/3-21G//3-21G geometry. The differences between those data and the present ones reflect almost exclusively the use of different sampling schemes rather than different geometries.

sented in Table II reproduce the quantum mechanical dipole moment components extremely accurately, since dipole moment constraints were used in the fitting procedures. The charges produced by a straightforward population analysis of the electronic wave function in the ZDO approximation[43] do not perform as well. For example, the ZDO charges from the INDO/S SDCI wave functions lead to classical dipole moments of 1.21 D (S_0), 1.67 D (^1L_b), and 4.92 D (^1L_a), respectively (Table II). These charges considerably underestimate the polarity of the ground state but predict the dipole moment of the excited ^1L_a state well. The INDO/S ZDO partial atomic charges for S_0 of 3-MeIn are mostly even smaller in absolute magnitude than those given by the electrostatic potential fitting procedures (Table II), which in turn are distinctly smaller than those obtained from the *ab initio* and MNDO calculations (Table I). The use of ZDO (or Mulliken) charges to simulate spectral shifts of indoles induced by external perturbations does not seem justifiable.[44]

CONCLUDING REMARKS

We have extended the computation of electrostatic potentials and potential derived partial atomic charges to correlated wave functions obtained from *ab initio* or semiempirical procedures and representing ground or electronically excited states. Although electrostatic interactions may not play the same important role in discussions of excited state processes as they do in the analysis of ground state conformation and reactivity, we believe that electrostatic potentials and potential derived charges will become very useful in elucidating and simulating excited state processes as well. As one example of the kind of problem we have in mind, detailed knowledge of excited state wave functions and/or partial atomic charges are a required prerequisite for molecular simulations of solvent effects on photoinduced charge transfer reactions.

For the ground state of 3-MeIn, the effects of correlation were small in the electrostatic potentials or potential derived charges. This may be related to the fact that the electrostatic potential is the expectation value of a one-electron operator and thus should be well approximated by an HF type wave function. Also, 3-MeIn features a rather conventional bonding pattern and the absence of polar groups, lone pairs, etc. Larger differences in electrostatic potentials (and hence fitted charges) between independent particle and correlated wave functions would certainly be expected in derivatives such as tryptophane, when the amino acid is in its zwitterionic form. Whereas ground state charge distributions can be obtained routinely by *ab initio* methods for rather large molecules of even low symmetry with high quality basis sets, i.e., well-defined, reliable charge sets can be developed, the situation is quite different in the context of electronically excited states. We feel that semiempirical methods present an advantage for such cases although calibration of the electrostatic potentials and potential derived charges may be necessary. In 3-MeIn, the semiempirical methods tend to produce charges smaller in magnitude than those given by a high level *ab initio* calculation, even when the computed dipole moments are virtually identical. The optimal scale factor between our INDO/S and *ab initio* 6-31G* potential derived charges at the HF level for the ground state of 3-MeIn is 1.44, extremely close to the recommended scale factor of 1.42[13] between the corresponding MNDO and 6-31G* potential derived charges or the value of 1.36 actually obtained for 3-MeIn in this work. We caution that this striking similarity in optimal, least-squares, scale factors between the semiempirical (INDO/S, MNDO) and *ab initio* (HF/6-31G*) potential derived charges does not necessarily imply that properties computed from molecular dynamics simulations with scaled INDO/S or MNDO charges will be closely similar to those obtained from the HF/6-31G* charges. For example, we find from a recently proposed dielectric formula based on linear responses theory[8] that the electrostatic contribution to the solvation energies given by the INDO/S and MNDO charges (both scaled) for 3-MeIn in water differ significantly ($\sim$6.5 kcal/mol) from each other but also that the INDO/S and HF/6-31G* values agree to within 1.5 kcal/mol.[45]

A new sampling procedure has been introduced to help alleviate the substantial rotational variance encountered in commonly used codes. We stress again the need for careful sampling, not only in the case of conformational analysis[19,20] but also when the basic aim is to produce charges for use in molecular dynamics simulations. Most molecules of biological interest possess little or no symmetry whatsoever in their equilibrium conformation and there is thus no obvious, preferred global coordinate system to choose for the calculations. Inspection of potential derived charges in such molecules based on a single calculation provides no information regarding to what extent the results obtained are, in fact, rotationally invariant. Caution should be exercised with the standard electronic structure packages that automatically enforce specific defaults in the sampling procedures.

Clearly, other parameterizations, computational methods, or levels will produce alternative sets of partial atomic charges relative to the ones presented here, even if the computed quantum me-

chanical dipole moments are similar (cf. HF values for S_0 of 3-MeIn in Tables I and II). Whereas the present method is based on creating an optimal fit to the molecular electrostatic potential, the approach taken by Dinur and Hagler, for example, is to define atomic multipoles based on electrostatic force considerations.[46,47] Their scheme appears very promising, although it has currently only been fully analyzed and implemented for planar molecules. There are no quantum mechanical rules or direct experimental data available for deciding which, if any, charge set is correct. In molecular dynamics simulations, the magnitudes of the partial atomic charges and the molecular dipole moment dominate the solute–solvent interactions, and it may be possible through simulations to provide a means of establishing optimal sets of ground and excited charges.[48,49]

Finally, the present charges all refer to an idealized vacuum situation. Solute–solvent polarization will alter the electronic distribution of the solute and hence the atomic charges for use in condensed phase molecular dynamics simulations ought to be different from the vacuum charges. We are also working on incorporating polarization into such simulations and will report on our progress presently.[45]

Financial support for this work was provided by a Biomedical Research Support grant (PHS RR 07058-25) and grants from the Rutgers Research Council and the David and Johanna Busch Memorial Fund. The final stages of the research was also supported by the National Science Foundation (DMB 91-05208). The authors thank the Pittsburgh Supercomputer Center for a generous grant of computer time.

References

1. E. Scrocco and J. Tomasi, *Top. Curr. Chem.*, **42,** 95 (1973); *Adv. Quantum Chem.*, **11,** 115 (1978).
2. P. Politzer and D.G. Truhlar, Eds., *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Plenum Press, New York, 1981.
3. F.A. Momany, *J. Phys. Chem.*, **82,** 592 (1978).
4. S.R. Cox and D.E. Williams, *J. Comp. Chem.*, **2,** 304 (1981).
5. U.C. Singh and P.A. Kollman, *J. Comp. Chem.*, **5,** 129 (1984).
6. J.T. Blair, K. Krogh-Jespersen, and R.M. Levy, *J. Am. Chem. Soc.*, **111,** 6948 (1989).
7. R.M. Levy, D.B. Kitchen, J.T. Blair, and K. Krogh-Jespersen, *J. Phys. Chem.*, **94,** 4470 (1990).
8. R.M. Levy, J.D. Westbrook, D.B. Kitchen, and K. Krogh-Jespersen, *J. Phys. Chem.*, **95,** 6756 (1991).
9. J.E. Ridley and M.C. Zerner, *Theor. Chim. Acta*, **32,** 111 (1973).
10. K. Krogh-Jespersen and M.A. Ratner, *J. Chem. Phys.*, **65,** 1305 (1976); *Theor. Chim. Acta*, **47,** 283 (1978).
11. R. McWeeny, *Phys. Rev.*, **126,** 1028 (1962); R. McWeeny, *Methods of Molecular Quantum Mechanics*, 2nd ed., Academic Press, New York, 1978, p. 115.
12. L.E. Chirlian and M.M. Francl, *J. Comp. Chem.*, **8,** 894 (1987).
13. B.H. Besler, K.M. Merz Jr., and P.A. Kollman, *J. Comp. Chem.*, **11,** 431 (1990).
14. M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegel, K. Raghavachari, M. Robb, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin, L.R. Kahn, J.J.P. Stewart, S. Topiol, and J.A. Pople, Gaussian 90, Gaussian, Inc., Pittsburgh, PA, 1990.
15. S.F. Boys, C.M. Reeves, and I. Shavitt, *Nature*, **178,** 1207 (1956); C.M. Reeves, PhD Thesis, Cambridge University, Cambridge, UK, 1957.
16. (a) C.M. Reeves, *Comm. ACM*, **9,** 276 (1966); (b) B.T. Sutcliffe, *J. Chem. Phys.*, **45,** 235 (1966); (c) G.H.F. Diercksen and B.T. Sutcliffe, *Theor. Chim. Acta*, **34,** 105 (1974); (d) I. Shavitt, in *Methods of Electronic Structure Theory*, H.F. Schaefer, Ed., Plenum Press, New York, 1977.
17. L.E. Chirlian and M.M. Francl, *QCPE Bull.*, **7,** 524 (1987).
18. K.M. Merz and B.H. Besler, *QCPE Bull.*, **10,** 15 (1990).
19. C.M. Breneman and K.B. Wiberg, *J. Comp. Chem.*, **11,** 361 (1990).
20. R.J. Woods, M. Khalil, W. Pell, S.H. Moffat, and V.H. Smith Jr., *J. Comp. Chem.*, **11,** 297 (1990).
21. J.J.P. Stewart, *QCPE Bull.*, **6,** 391 (1986).
22. R. Weinstock, *Calculus of Variations*, Dover, New York, 1974.
23. C. Giessner-Prettre and A. Pullman, *Theor. Chim. Acta*, **25,** 83 (1972).
24. A.J. Duben, *Theor. Chim. Acta*, **59,** 81 (1981).
25. A. Chung-Phillips, *J. Comp. Chem.*, **10,** 17 (1989).
26. J.C. Culberson and M.C. Zerner, *Chem. Phys. Lett.*, **122,** 436 (1985).
27. P.O. Lowdin, *J. Chem. Phys.*, **56,** 365 (1970).
28. W.J. Hehre, R.F. Stewart, and J.A. Pople, *J. Chem. Phys.*, **51,** 2657 (1969).
29. M.W. Schmidt, J.A. Boatz, K.K. Baldridge, S. Koseki, M.S. Gordon, S.T. Elbert, and B. Lam, *QCPE Bull.*, **7,** 115 (1987).
30. J.D. Westbrook and K. Krogh-Jespersen, ESPPAC, Rutgers University, New Brunswick, NJ, 1989.
31. A.L. McClellan, *Tables of Experimental Dipole Moments*, Freeman, London, 1963.
32. P. Politzer and K.C. Daiker, *Chem. Phys. Lett.*, **34,** 294 (1975).
33. P. Politzer and K.C. Daiker, in *Force Concept in Chemistry*, B.M. Deb, Ed., Van Nostrand Reinhold, New York, 1981.
34. R.S. Mulliken, *J. Chem. Phys.*, **23,** 1833 (1955).
35. J.R. Beecham and L. Brand, *Annu. Rev. Biochem.*, **54,** 43 (1985).
36. P.R. Callis, *J. Chem. Phys.*, **95,** 4230 (1991).
37. J.R. Platt, *J. Chem. Phys.*, **85,** 1377 (1986).
38. H. Lami and N. Glasser, *J. Chem. Phys.*, **84,** 597 (1986); S.R. Meech, D. Phillips, and A.G. Lee, *Chem. Phys.*, **80,** 317 (1983); M. Sun and P.S. Song, *Photochem. Photobiol.*, **25,** 3 (1977).
39. C.-T. Chang, C.-Y. Wu, A.R. Muirhead, and J.R. Lombardi, *Photochem. Photobiol.*, **19,** 347 (1974).
40. J.R. Cable, *J. Chem. Phys.*, **92,** 1627 (1990).
41. A.A. Rehms and P.R. Callis, *Chem. Phys. Lett.*, **140,** 83 (1987).
42. F.J. Luque, F. Illas, and M. Orozco, *J. Comp. Chem.*, **11,** 416 (1990).

43. J.A. Pople and D.L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill, New York, 1970.
44. P. Ilich, C. Haydock, and F.G. Prendergast, *Chem. Phys. Lett.*, **158,** 129 (1989); P. Ilich, P.H. Axelsen, and F.G. Prendergast, *Biophys. Chem.*, **29,** 341 (1988).
45. J.D. Westbrook, R.M. Levy, and K. Krogh-Jespersen, to be submitted.
46. U. Dinur and A.T. Hagler, *J. Chem. Phys.*, **91,** 2959 (1989); *J. Chem. Phys.*, **91,** 2949 (1989).
47. U. Dinur, *J. Comp. Chem.*, **12,** 469 (1991).
48. F.H. Stillinger and A. Rahman, *J. Chem. Phys.*, **60,** 1545 (1974); *J. Chem. Phys.*, **68,** 666 (1978).
49. W.L. Jorgensen, J.D. Madura, and C.J. Swensen, *J. Am. Chem. Soc.*, **106,** 6638 (1984); W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *J. Chem. Phys.*, **79,** 926 (1983).