Analysis of a Large Data Base of Electrostatic Potential Derived Atomic Charges

Kenneth M. Merz Jr.

The Department Of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

Received 16 October 1991; accepted 7 February 1992

A large data base of 6-31G*, MNDO, AM1, and PM3 electrostatic potential (ESP) derived point charges of amino acids and monosaccharides is analyzed. We find that MNDO correlates well with 6-31G* ESP derived point charges, while AM1 and PM3 do so quite poorly. Furthermore, scaling MNDO ESP derived point charges enhances the ability of MNDO to reproduce 6-31G* results. We used our data base to attempt to derive a 6-31G* transferable charge model at an atom-by-atom level. We find that it is simple to derive a transferable model for monosaccharides, but for the amino acids statistical difficulties make this a less attractive approach. The transferable charge model for the monosaccharides is slightly better than MNDO, but scaled MNDO charges perform significantly better than the transferable model. We also carried out a QMD simulation on the alanine dipeptide to assess the fluctuations that would be expected in atomic point charges during the course of an MD simulation. Relatively large charge fluctuations are observed and their impact on molecular simulation is addressed.

INTRODUCTION

Electrostatic interactions are very important mediators of the structure and function of macromolecular systems.[1-3] Therefore, the accurate modeling of these interactions is an important contemporary goal of theoretical chemistry and biophysics. The usual (and simplest) representation of the electrostatic properties of a system is through atom centered point charges (monopole approximation) and many techniques have been proposed over the years to accurately produce a molecular charge distribution. These range from charges determined using Mulliken population analysis,[4] electrostatic potential (ESP) fitting using *ab initio*[5,6] and semiempirical techniques,[7-11] empirical approaches (i.e., fitting charges to reproduce experimental data, etc.),[12,13] and the determination of charges from experiment.[14-16] In recent years techniques aimed at reproducing the electrostatic properties of a molecule via a multipolar expansion have appeared.[17-19]

We recently developed an approach to determining atomic point charges using the semiempirical molecular orbital method MNDO[20,21] in conjunction with the ESP fitting technique.[7] It was found that this approach was able to reasonably reproduce molecular electrostatic potentials as well as point charges relative to 6-31G*[4] *ab initio* calculations.

However, several key issues were not thoroughly investigated in our previous work on the semiempirical ESP technique. First, we have not addressed the orientation dependence[22] of the Connolly surface algorithm.[23] For example, Breneman and Wiberg[22] found that the surface used by Chirlian and Francl[24] in their CHELP program had a rather severe dependence upon the orientation of a molecule in 3D space. Furthermore, they found that CHELP was unable to smoothly map out the changes in molecular charge as the conformation of a molecule was changed. They suggested a modification to CHELP called CHELPG which mitigated these problems. In the present article we investigated the orientation dependence of the Connolly algorithm. In a related issue we also investigated the convergence of charges vs. the density of points considered.

Second, in our previous effort we looked at a limited number of small molecules to determine how well semiempirical techniques perform relative to *ab initio* calculations. Given the much larger data base of amino acid and monosaccharide ESP derived atomic point charges we now have at our disposal we should be able to better define how well the semiempirical techniques MNDO, AM1,[25] and PM3[26-28] do when they are employed to determine atomic point charges with the ESP fitting technique.

Third, our large data base allows us to determine if we can derive an atom-by-atom[19] (meaning average over atoms in similar molecular environments) 6-31G* transferable charge model that is capable

CCC 0192-8651/92/060749-19$04.00

of giving a more accurate representation of the electrostatic properties of a molecule than does the semiempirical ESP techniques.

Finally, the amount by which the point charges for a molecular system can change due to rotations about single bonds has been addressed by several authors[22,29,30] and has been found to be significant in many cases. To further address this we developed a quantum molecular dynamics (QMD) technique that allows us to evaluate point charges at a number of molecular structures along a QMD trajectory. Previous authors examined several selected conformations of alanine dipeptide[29] and how atomic point charges vary along the rotation profile of the C—N bond in formamide.[22,30] The question we are asking is a little different than that addressed by these authors[22,29,30] in that we want to know how much atomic point charges fluctuate during the course of a 300K molecular dynamics (MD) simulation. This sort of analysis will give us insight into the possible errors made when using fixed atomic point charges during an MD simulation.

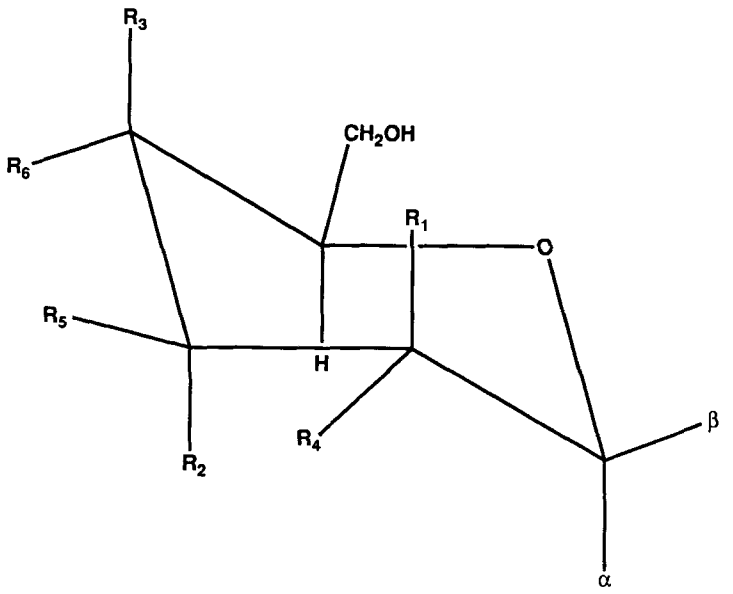
In the present article we address all of the aforementioned issues using *ab initio* 6-31G* and semiempirical MNDO, PM3, and AM1 ESP calculations. The discussion of our results is given in four sections: The first discusses the orientation and ESP point density dependences of ESP derived atomic point charges. This is followed by a discussion of a general analysis of the data base, which again reevaluates the quality of semiempirical ESP derived atomic point charges. We then discuss atom-by-atom[19] transferable atomic point charges and compare the quality of this model with semiempirical point charges. Finally, we address the charge fluctuations present during an MD simulation.

COMPUTATIONAL PROCEDURE

A modified version[31] of Gaussian 88[32] was used to determine the 6-31G* ESP derived point charges. For the smaller amino acids (Gly, Ala, Val, Ile, Ser, Thr, Cys, Met, Asn, Gln, Asp, Glu, His, and Pro) G80 UCSF was used.[33] The Connolly surface routine was used in Gaussian 88 and G80 UCSF to generate four layers of points 0.2 Å apart with the first one being 1.4 times the van der Waals radii of the constituent atoms of a molecule. The density of points in each layer was one point per $Å^2$ of surface area. The direct SCF procedure was used in the Gaussian 88 calculations[34] and disk bound SCF procedures were used for the G80 calculations. The structures studied are given in Figure 1. The dipeptide side-chain geometries were taken from experimental neutron diffraction data[35] whenever possible; otherwise bond lengths and angles were taken from a compilation of these values.[36] The main-chain geometry was identical in all cases and the bond distances and angles were taken from the compilation.[36] For the monosaccharides the neutron structure of α-D-glucose was used to generate all the other saccharide geometries.[37] The geometries used are given in MOPAC 5.0[38] Cartesian coordinate formate as supplementary material. The molecular orientation as well as the ESP surface generated by Gaussian 88 and G80 UCSF was then used to determine the semiempirical ESP derived point charges using the MNDO, AM1, and PM3 Hamiltonians. This eliminated any surface dependence in the ESP derived charges that were used in the comparison of 6-31G* and semiempirical atomic point charges. The STO-3G basis set was used in the deorthogonalization procedure rather than the STO-6G basis set.[7] We have found that this speeds the ESP procedure and does not have a significant effect on the ESP potential nor on the atomic point charges. The resulting charges were studied using linear regression analysis. It should be noted that an atomic point charge model cannot give an exact fit to an electrostatic potential; thus when considering atomic charges derived using an atom centered fitting procedure the goodness of fit is considered by evaluating the root mean square (rms) fits.[5] For all cases considered here the rms error is significantly less than 1 kcal/mol, which is indicative of a good fit between the quantum mechanically derived ESP and that obtained using atomic point charges.

To sample a statistically significant number of molecular conformations we carried out QMD simulations using the MNDO Hamiltonian. The QMD method involves using the forces determined from the quantum mechanical potential to drive the atoms classically using Newton's equations of motion. Thus, the forces and potential energies are determined quantum mechanically and conformation space is sampled using classical mechanics. A number of other authors have described QMD simulations using both density functional techniques[39,40] and semiempirical techniques.[41,42]

The system used in our QMD simulation is the alanine dipeptide (see Fig. 1). This was a prudent model because it is conformationally flexible and because it is small enough to carry out expensive QMD simulations on. The calculations were carried out as follows: The system was first fully minimized into an α-helical conformation (Φ, $\Psi \approx -60°$) and the initial velocities were taken from a Boltzmann distribution. The initial temperature was chosen to be 300 K. A constant energy simulation was done and analytic gradients[43] were used because they were found to improve the energy conservation.[31] The system was equilibrated for 25 ps which was followed by 50 ps of data collection.



R1 = R2 = R3 = H; R4 = R5 = R6 = OH; α = OH; β = H; α-D-Glucose
R1 = R2 = R3 = H; R4 = R5 = R6 = OH; α = H; β = OH; β-D-Glucose
R2 = R3 = R4 = H; R1 = R5 = R6 = OH; α = OH; β = H; α-D-Mannose
R2 = R3 = R4 = H; R1 = R5 = R6 = OH; α = H; β = OH; β-D-Mannose
R1 = R5 = R6 = H; R2 = R3 = R4 = OH; α = OH; β = H; α-D-Gulose
R1 = R5 = R6 = H; R2 = R3 = R4 = OH; α = H; β = OH; β-D-Gulose
R4 = R5 = R6 = H; R1 = R2 = R3 = OH; α = OH; β = H; α-D-Idose
R4 = R5 = R6 = H; R1 = R2 = R3 = OH; α = H; β = OH; β-D-Idose
R1 = R3 = R5 = H; R2 = R4 = R6 = OH; α = OH; β = H; α-D-Allose
R1 = R3 = R5 = H; R2 = R4 = R6 = OH; α = H; β = OH; β-D-Allose
R3 = R4 = R5 = H; R1 = R2 = R6 = OH; α = OH; β = H; α-D-Altrose
R3 = R4 = R5 = H; R1 = R2 = R6 = OH; α = H; β = OH; β-D-Altrose
R1 = R2 = R6 = H; R4 = R3 = R5 = OH; α = OH; β = H; α-D-Galactose
R1 = R2 = R6 = H; R4 = R3 = R5 = OH; α = H; β = OH; β-D-Galactose
R2 = R4 = R6 = H; R1 = R3 = R6 = OH; α = OH; β = H; α-D-Talose
R2 = R4 = R6 = H; R1 = R3 = R6 = OH; α = H; β = OH; β-D-Talose

Figure 1. Molecules examined in this study.

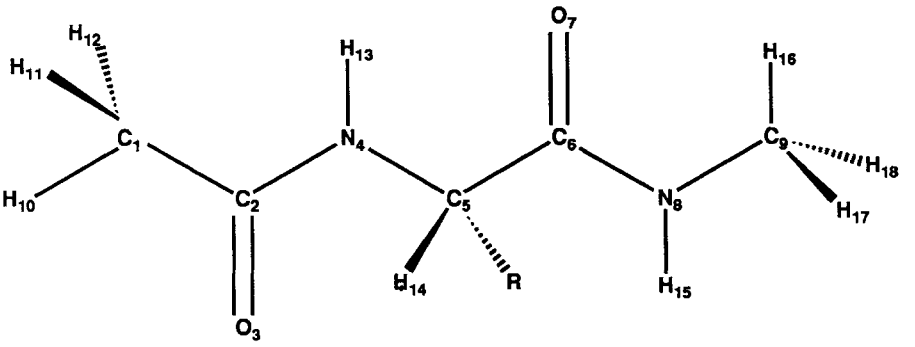
The various quantities to be averaged were saved every 25 timesteps and a timestep of 1 fs was used. This simulation protocol results in 2K sets of data points for later analysis. A modified version of MOPAC 5.0 ESP[38] was used to carry out the simulations and MONAL 1.0[31] was used to analyze the trajectories.

RESULTS AND DISCUSSION

Dependence of Charges on the Molecular Orientation

To address the dependence of atomic point charges on molecular orientation we carried out 10 MNDO ESP calculations at a fixed extended conformation where we altered the molecular orientation in 3D space by 30, 45, and 90° along the x, y, and z axes. The results of these calculations are given in Table I. Given the small standard deviation it is apparent that orientational effects are only a small error in the determination of ESP charges. Thus, the Connolly surface generating algorithm, under the conditions we use it, is not very dependent upon orientation in 3D space.

A related question is the convergence of the ESP charges that are obtained using our procedure. To address this we carried out seven separate ESP evaluations using various densities of points (1, 5, 10, 15, 20, 25, and 30 points per Å^2). This resulted in the evaluation of the electrostatic poten-



R = H (Gly), CH_3 (Ala), $CH(CH_3)_2$ (Val), $CH_2CH(CH_3)_2$ (Leu), $CH(CH_2CH_3)(CH_3)$ (Ile), $CH_2C_6H_5$ (Phe), $CH_2C_6H_4$-p-OH (Tyr), $CH_2C_8H_6N$ (Trp), CH_2OH (Ser), $CH(OH)CH_3$) (Thr), CH_2SH (Cys), $CH_2CH_2SCH_3$ (Met), $CH_2C(=O)NH_2$ (Asn), $CH_2CH_2C(=O)NH_2$ (Gln), CH_2COO^- (Asp), CH2CH2COO$^-$ (Glu), $CH_2CH_2CH_2CH_2NH_3^+$ (Lys), $CH_2CH_2CH_2NHC(=NH_2^+)NH_2$ (Arg), $CH_2C_3H_3N_2$ (His two forms H_ε and H_δ), $CH_2C_3H_4N_2$ (His protonated)

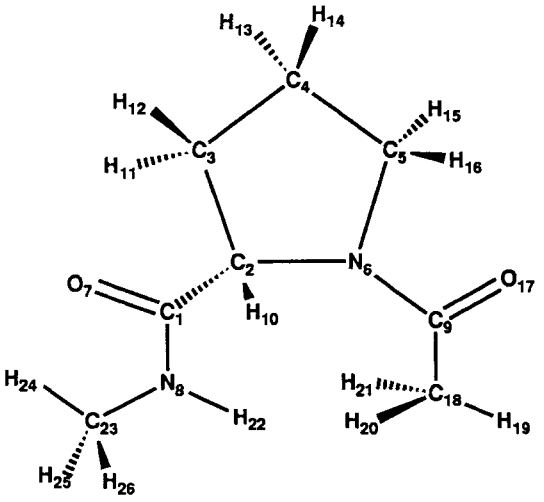


Figure 1. (continued)

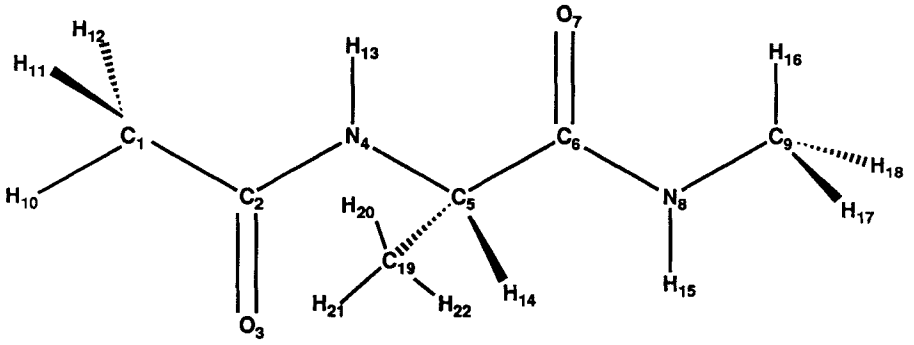
tial at 978, 5066, 10,399, 14,612, 16,872, 18,087, and 18,601 points, respectively. For the consideration of further points the number of layers has to be altered. The results are given in Table II. From the low standard deviation in the charges and dipole moment it is clear that the low point density is almost as good a representation as is the high density. Furthermore, all charge sets correlate well with one another and all have correlations coefficients of greater than 0.99.

General Analysis

The data base consists of 1007 individual observations (i.e., charges) with 384 of these observations coming from the monosaccharides and 623 from the amino acids. Examples of the ESP derived point charges are given in Tables III and IV. Table III contains the point charges for the tryptophan dipeptide, while Table IV contains those for α- and β-D-idose. The remainder of the values will be reported in subsequent publications describing force field developments.[31,44] The database consists of 6-31G*, MNDO, AM1, and PM3 ESP derived point charges. For the AM1 case compounds containing sulfur (i.e., Cys and Met) are not considered due to the lack of AM1 sulfur parameters. The range of the charges are: 6-31G*, −1.13 to 1.15; MNDO, −0.99 to 0.98; AM1, −1.02 to 1.25; PM3, −1.31 to 1.76. MNDO charges are uniformly less than the 6-31G* atomic point charges, while for AM1 and PM3 the charges are usually less than the 6-31G* charges with a few exceptions. In general, we found that the semiempirical ESP point charges are less than the 6-31G* *ab initio* values. This is consistent with our previous observations.[7]

Figure 2 gives the fit of the MNDO ESP derived point charges to the corresponding 6-31G* values

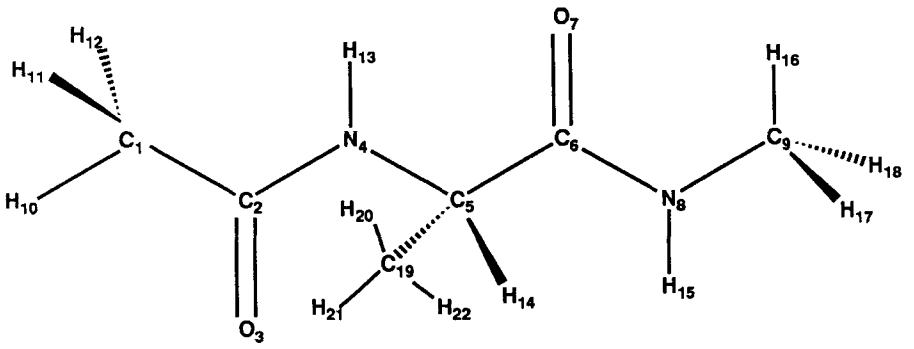
Table I. Orientation dependence of ESP derived point charges.



Atom label	MNDO (AVE)	MNDO (σ)
C−1	−0.3123	0.0099
C−2	0.6292	0.0053
O−3	−0.4913	0.0014
N−4	−0.6551	0.0076
C−5	0.2391	0.0242
C−6	0.4199	0.0113
O−7	−0.4350	0.0018
N−8	−0.4571	0.0090
C−9	−0.0603	0.0103
H−10	0.0888	0.0026
H−11	0.0834	0.0024
H−12	0.0887	0.0025
H−13	0.3337	0.0024
H−14	0.0180	0.0065
H−15	0.3036	0.0028
H−16	0.0692	0.0028
H−17	0.0733	0.0028
H−18	0.0816	0.0028
C−19	−0.1851	0.0143
H−20	0.0894	0.0038
H−21	0.0192	0.0041
H−22	0.0591	0.0035
μ (Debye)	2.66	0.003

using $y = mx$. We also examined the other semiempirical methods and the correlation coefficients for these fits are given in Table V. PM3 does very poorly and gives a correlation coefficient of only 0.7. AM1 performs a little better, but its correlation coefficient of 0.81 is also low. Finally, Figure 2 indicates that MNDO performs much better than PM3 and AM1 and this is borne out by the correlation coefficient of 0.96. All of the semiempirical ESP charges can be scaled to improve their correlation with the 6-31G* values, but it only makes sense to do this for MNDO. For MNDO the appropriate scale factors are 1.29 (all molecules), 1.35 (neutral molecules only), and 1.08 (charged molecules only). We can also separately scale for the monosaccharides (1.47) and amino acids (1.26). The scale factors reported here are to be compared to the 1.42 value reported previously for small molecules.[7] Suggestions regarding which scale factor to use will be given in the following section. Charged species are difficult to handle when determining a scale factor since it is impossible to keep an integral charge of +1 or −1 using $y = mx$. By fitting to $y = mx + b$ we can use b to correct the line and bring it approximately through the origin. This results in maintaining the integral charge. However, we find (see below) that scaling point charges for a charged molecule does not improve the fit between a 6-31G* and MNDO electrostatic potential. Hence, it is recommended to use "raw" MNDO charges for charged molecules; however, more research on a larger data base of charged molecules is needed before a firm conclusion can be made.[45] Finally, we should point out that it is **not** clear that scaling is going to guarantee improvement of the charge model, so testing the

Table II. Surface density dependence of ESP derived point charges.



Atom label	MNDO (AVE)	MNDO (σ)
C−1	−0.3072	0.0028
C−2	0.6321	0.0077
O−3	−0.4920	0.0023
N−4	−0.6591	0.0157
C−5	0.2362	0.0258
C−6	0.4197	0.0164
O−7	−0.4356	0.0033
N−8	−0.4546	0.0099
C−9	−0.0599	0.0056
H−10	0.0872	0.0009
H−11	0.0814	0.0007
H−12	0.0868	0.0007
H−13	0.3363	0.0031
H−14	0.0188	0.0049
H−15	0.3034	0.0027
H−16	0.0684	0.0009
H−17	0.0731	0.0015
H−18	0.0814	0.0013
C−19	−0.1800	0.0123
H−20	0.0878	0.0025
H−21	0.0179	0.0025
H−22	0.0579	0.0294
μ (Debye)	2.6662	0.0028

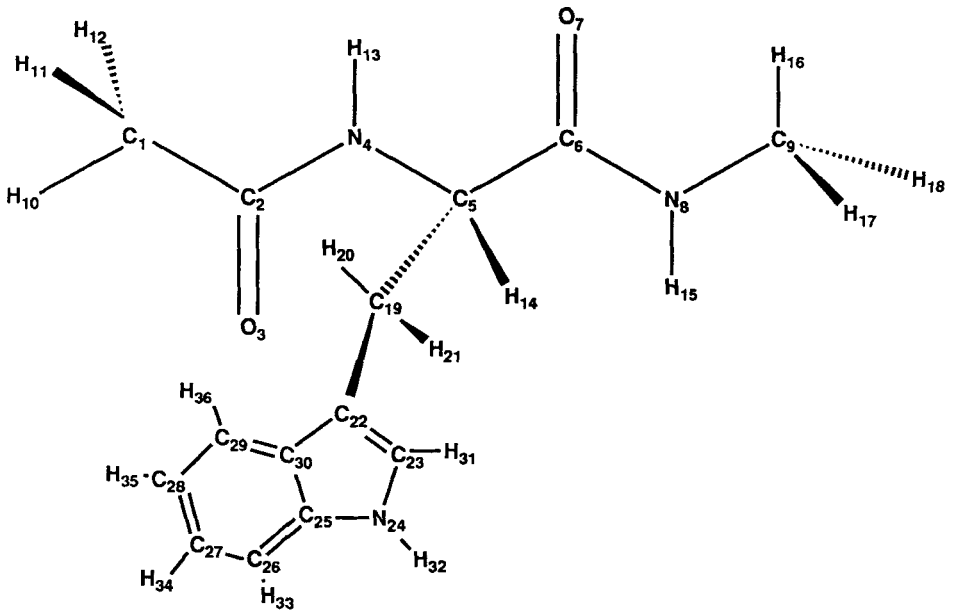
scaled charges vs. calculated dipole moments, electrostatic potentials, interaction energies, etc. is suggested to justify their use.

The fits of MNDO ESP point charges to 6-31G* for the monosaccharides and for the amino acids gives correlation coefficients of 0.99 and 0.97, respectively. The scale factor for the former is 1.47 while for the latter it is 1.26. Charged amino acids give a correlation coefficient of 0.92 and a scale factor of 1.08. The lower correlation coefficient for the amino acids is due mostly to the nitrogen atoms (see below), where MNDO performs poorly in reproducing 6-31G* ESP point charges.

Transferability of Charges

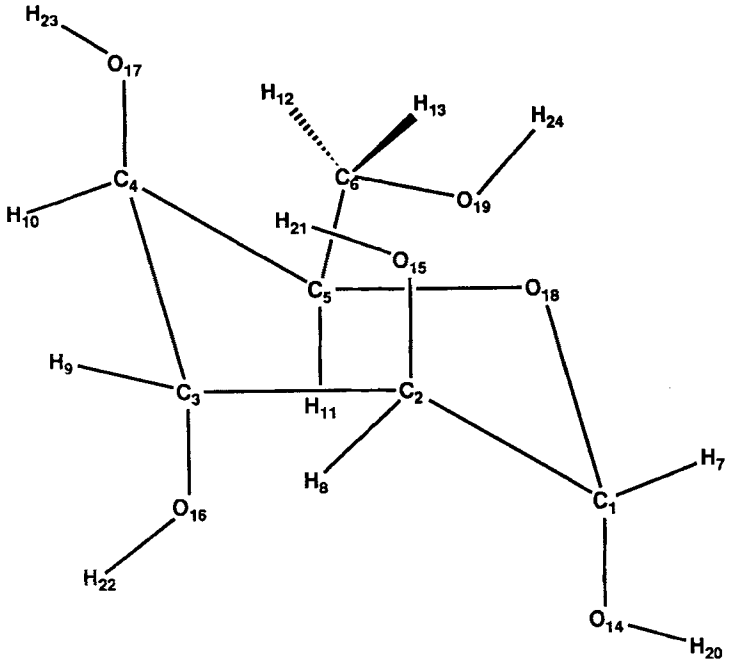
Given that we have a large data base of charges with a wide range of functionalities we should be able to address how transferable ESP derived point charges might be. There are a number of possible transferable charge models. In the new force fields being developed using our data base charges the transferable unit will be a residue. For a polysaccharide this is a monosaccharide unit and for a protein this is an amino acid unit. These models will be evaluated and described elsehwere.[31,44] An alternative transferable charge model is to average over values obtained from atom centers in a similar chemical environment.[19,46,47] In these efforts authors[19,46,47] have demonstrated that a high degree of transferability of a functional group (i.e., a CH_2 unit) is possible providing the adjacent groups are identical. Hence, a CH_2 group with two adjacent CH_2 groups is transferable into another system if the same set of functional groups are present. If another functional group pattern is present the

Table III. Atomic point charges obtained for tryptophan.



Atom label	6-31G*	MNDO	AM1	PM3
C−1	−0.43710	−0.26100	−0.49640	−0.24800
C−2	0.78192	0.58690	0.41570	0.25990
O−3	−0.65996	−0.48340	−0.37050	−0.39830
N−4	−0.43904	−0.51420	−0.49520	−0.16270
C−5	−0.26095	−0.10160	−0.08580	−0.03060
C−6	0.75901	0.52720	0.25980	0.15200
O−7	−0.63066	−0.46490	−0.33670	−0.37930
N−8	−0.48095	−0.46600	−0.34460	−0.02760
C−9	−0.25600	−0.04870	−0.34240	−0.18990
H−10	0.13380	0.08310	0.15940	0.08550
H−11	0.10810	0.06690	0.14880	0.07860
H−12	0.11116	0.06870	0.15350	0.08480
H−13	0.27408	0.30930	0.35150	0.20870
H−14	0.15104	0.10890	0.16100	0.10290
H−15	0.35770	0.31420	0.33170	0.18190
H−16	0.13129	0.06920	0.14930	0.07850
H−17	0.13043	0.07000	0.15140	0.08750
H−18	0.13676	0.07350	0.15640	0.09220
C−19	−0.14113	−0.09780	−0.33910	−0.22840
H−20	0.17713	0.12610	0.22420	0.17650
H−21	0.04515	0.04670	0.13790	0.08360
C−22	−0.09484	−0.14130	−0.07980	−0.05480
C−23	−0.06601	0.00810	−0.11860	−0.35750
N−24	−0.59124	−0.48250	−0.49580	0.00760
C−25	0.35773	0.24320	0.25500	0.10160
C−26	−0.36201	−0.27380	−0.37230	−0.33820
C−27	−0.08913	−0.02890	−0.09010	−0.07780
C−28	−0.25556	−0.18910	−0.26170	−0.21580
C−29	−0.14771	−0.03910	−0.11940	−0.11820
C−30	0.00401	−0.09320	−0.10650	−0.09580
H−31	0.23835	0.18070	0.26920	0.29510
H−32	0.39979	0.35500	0.39830	0.20220
H−33	0.18080	0.13940	0.21410	0.19450
H−34	0.13590	0.09300	0.16090	0.13760
H−35	0.15447	0.10930	0.17800	0.15290
H−36	0.14365	0.10600	0.17880	0.15880

Table IV. Atomic point charges obtained for α- and β-D-idose.

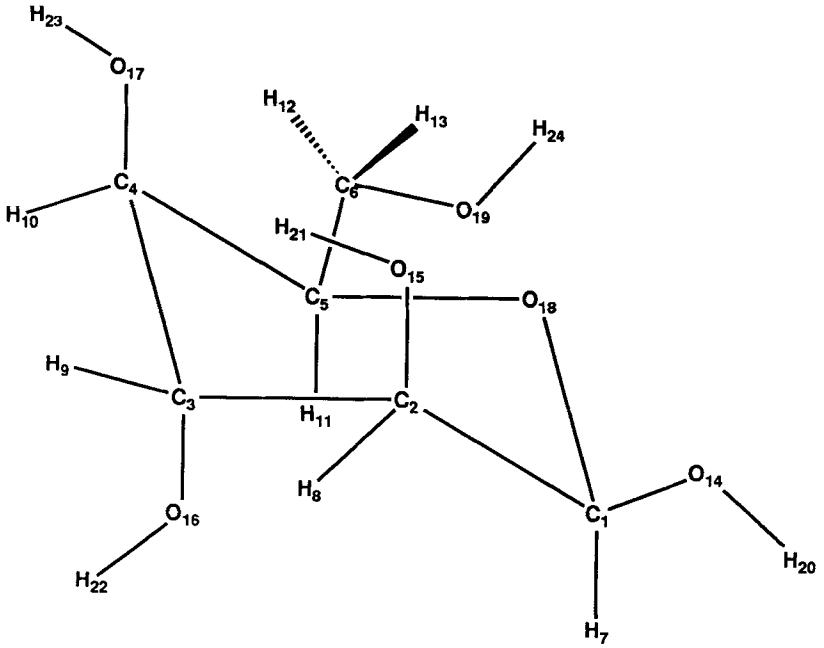


Atom label	6-31G*	MNDO	AM1	PM3
C—1	0.25787	0.21670	−0.18050	−0.15290
C—2	0.19921	0.14220	0.02440	0.11810
C—3	0.15664	0.08560	−0.22080	−0.22900
C—4	0.03645	0.03750	−0.02550	0.11370
C—5	0.28562	0.14730	0.01900	−0.05330
C—6	0.11030	0.11080	−0.21770	−0.10420
H—7	0.08257	0.07140	0.18810	0.14000
H—8	0.06069	0.04360	0.12040	0.06560
H—9	0.06660	0.05500	0.13400	0.08600
H—10	0.07521	0.06120	0.10820	0.04380
H—11	0.12732	0.10940	0.20020	0.18130
H—12	0.01850	0.00520	0.10000	0.05160
H—13	0.06258	0.05160	0.14000	0.09020
O—14	−0.65644	−0.47530	−0.30970	−0.27690
O—15	−0.63975	−0.45810	−0.30000	−0.27550
O—16	−0.69884	−0.47510	−0.30710	−0.25800
O—17	−0.60861	−0.42430	−0.26060	−0.21700
O—18	−0.39952	−0.31410	−0.11370	−0.08130
O—19	−0.69551	−0.47160	−0.33660	−0.30520
H—20	0.44782	0.31470	0.26800	0.23620
H—21	0.40756	0.28570	0.22890	0.19540
H—22	0.44923	0.30300	0.26650	0.23440
H—23	0.41188	0.28080	0.21450	0.17020
H—24	0.44261	0.29670	0.25980	0.22650

atomic point charges are less transferable. In the present article we will attempt to develop a transferable charge model by considering individual atoms (atom-by-atom model) and their charges determined via the ESP technique. The transferable models will then be tested against *ab initio* dipole moments and electrostatic potentials to assess how well they perform. We will first consider the C, H, N, and O classes to determine the variation in these charges as a whole and then consider the atom-by-atom breakdown. The latter will be done by considering the saccharides and amino acids separately since they each have features that are unique to their chemical architecture.

Table VI contains information regarding the average charges for C, O, N, and H for both MNDO

Table IV. (continued)



Atom label	6-31G*	MNDO	AM1	PM3
C−1	0.55129	0.26410	−0.07430	−0.03610
C−2	−0.06836	0.06360	−0.02890	0.01520
C−3	0.21049	0.02850	−0.25660	−0.22110
C−4	0.06151	0.12810	0.06060	0.16500
C−5	0.28583	0.06240	0.00660	−0.01010
C−6	0.25624	0.18000	−0.13470	−0.02960
H−7	0.07138	0.10010	0.21170	0.16000
H−8	0.10481	0.05730	0.11030	0.06890
H−9	0.07869	0.07230	0.14240	0.08890
H−10	0.06810	0.04930	0.09080	0.03250
H−11	0.07837	0.10370	0.17720	0.15130
H−12	−0.02693	−0.01220	0.06950	0.01670
H−13	0.02794	0.03840	0.12690	0.08190
O−14	−0.74133	−0.49190	−0.34310	−0.30700
O−15	−0.61988	−0.43890	−0.28180	−0.23550
O−16	−0.76624	−0.51010	−0.36660	−0.32650
O−17	−0.61106	−0.43180	−0.28530	−0.24910
O−18	−0.46342	−0.29870	−0.16500	−0.16210
O−19	−0.72051	−0.47870	−0.34500	−0.31260
H−20	0.47808	0.33510	0.30080	0.26520
H−21	0.42522	0.28670	0.22030	0.18280
H−22	0.46883	0.32020	0.28950	0.26020
H−23	0.40784	0.27570	0.21590	0.17700
H−24	0.44310	0.29700	0.25900	0.22430

and 6-31G*. We have 280 observations for C and the mean value for all 6-31G* carbons is 0.111 with a standard deviation of 0.424. The corresponding MNDO values are 0.070 and 0.282, respectively. We also determined how well MNDO correlates with the 6-31G* ESP charges for each of the atom classes. The correlation coefficients are also given in Table VI. For C we find that MNDO correlates reasonably well (0.93).

For O we have 150 individual observations, which yield a mean of −0.649 and a standard deviation of 0.109 for the 6-31G* values, while MNDO is −0.458 and 0.083, respectively. The correlation of MNDO with 6-31G* in this case is 0.91. For N we

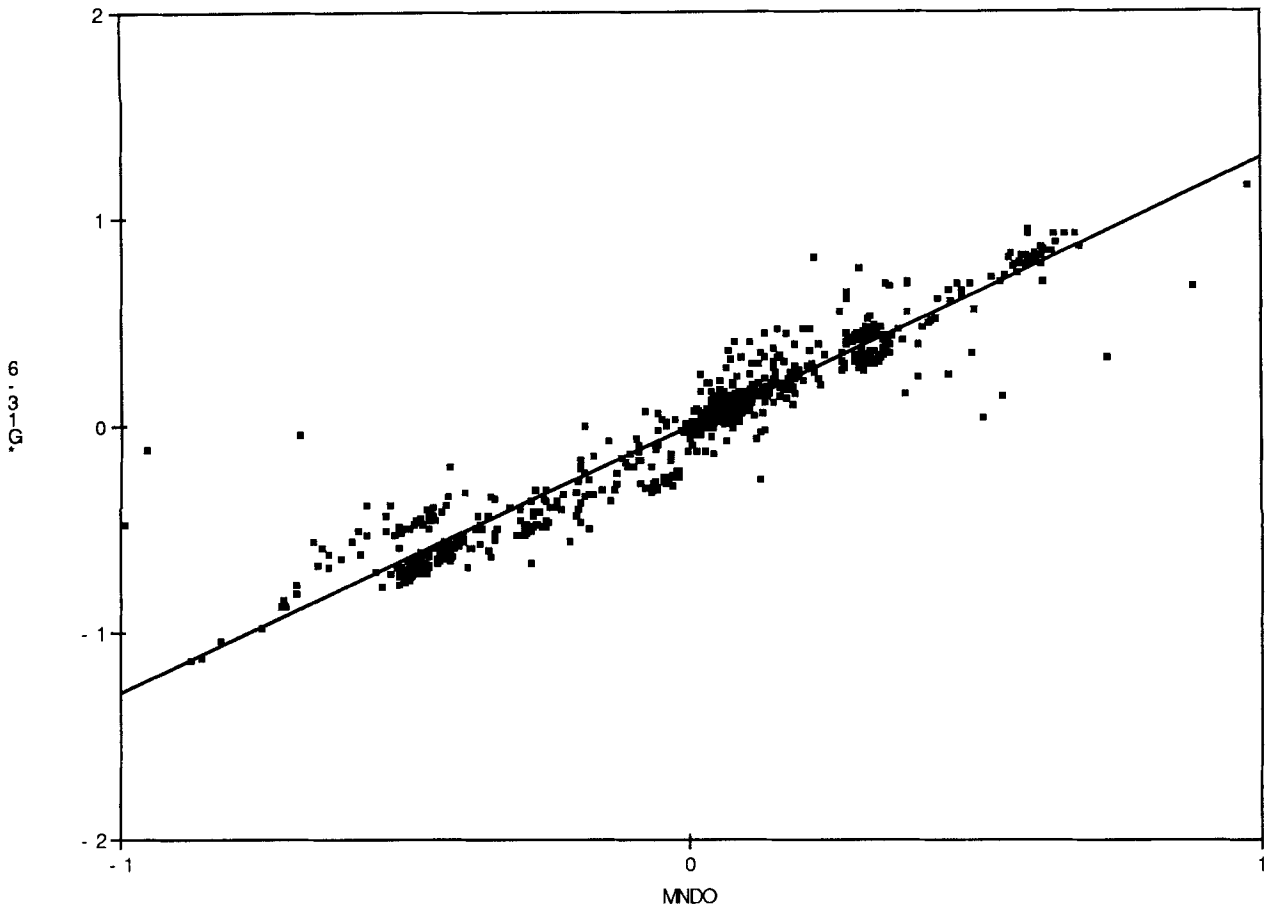


Figure 2. Correlation of MNDO with 6-31G* ESP derived atomic point charges.

have only 57 observations which give a mean of −0.521 and a standard deviation of 0.197 for 6-31G* charges. For MNDO the corresponding values are −0.519 and 0.171, respectively. The correlation between MNDO and 6-31G* for N is very poor being only 0.5. Finally, we have 518 observations of H atoms and we have lumped both the polar and nonpolar hydrogens together into one class. The average 6-31G* and MNDO values are 0.187 and 0.141 with standard deviations of 0.148 and 0.119, respectively. MNDO ESP charges correlate reasonably well with 6-31G* ones in this case with a correlation coefficient of 0.92. We also considered S atoms, but we only have two examples in the data base. For completeness, however, the average values and standard deviations are given in Table VI.

Table V. Correlation coefficients for the linear regression analysis of atomic point charges.

Method	6-31G*	MNDO	AM1	PM3
6-31G*	—	0.96	0.81	0.70
MNDO	0.96	—	0.87	0.76
AM1	0.81	0.87	—	0.88
PM3	0.70	0.76	0.88	—

Table VI. Averages and standard deviations for 6-31G* (MNDO) charge values as well as correlation coefficient between 6-31G* and MNDO charges.

Atom type	Average	SD	Correlation coefficient
C	0.111 (0.070)	0.424 (0.282)	0.93
H	0.187 (0.148)	0.141 (0.119)	0.92
O	−0.649 (−0.458)	0.109 (0.083)	0.91
N	−0.521 (−0.519)	0.197 (0.171)	0.50
S	−0.306 (−0.129)	0.042 (0.058)	—

Several important features arise from these considerations: (1), C, N, and H atoms vary quite widely at both the MNDO and 6-31G* level, while O has the smallest standard deviation relative to the others; (2) MNDO reproduces N atom centered points charges very poorly; (3) to develop a transferable charge model it is clear that further atom classes need to be considered. Another possible way in which to classify atom types is to consider the atomic hybridization. However, this classification scheme gives similarly large charge variations as does the preceding procedure. This can be confirmed by inspecting Table III where we have several types of sp^2 hybrid carbon atoms that have atomic charge values ranging from +0.78 to −0.36. This again indicates the need for more re-

fined atom classes. This is consistent with previous observations that indicated that clearly defined chemical environments or atoms of the same hybridization gave the most reasonable transferable models.[19,46,47]

We decided to first consider monosaccharides because of the homogeneity of these molecules. Hence, monosaccharides offer a clear-cut model to determine how well ESP techniques perform in the development of an atom-by-atom transferable charge model. This is not the case for amino acids due to the heterogeneity of the side-chain groups. Hence, we will use the monosaccharides as our test case to determine the efficacy of an atom-by-atom transferable charge model. We will then consider the amino acids in light of the monosaccharide results.

Monosaccharides

To test the ability to obtain an accurate transferable charge model we decided to consider two possibilities. In the first we considered only four subclasses: polar hydrogens, nonpolar hydrogens, oxygens (all oxygens), and finally carbon atoms. In the second class we averaged over each atom type for each monosaccharide (24 possible atom types in all). These two models allow us to consider the two possible extremes for a transferable charge model. Furthermore, we have found from extensive experimentation that the latter model will give us the best possible transferable charge model. The values obtained in this way along with some intermediate atom subclasses (e.g., ring vs. hydroxyl oxygens) are given in Table VII and MNDO values are included for completeness. The transferable model we are using is beyond other workers'[19,46,47] models that we will call the "functional group" model. In their model atoms in similar chemical environments (hybridization) are averaged together, but in ours we have considered each atom by itself regardless if it has a partner that has a similar chemical environment. From tests we have found that our model is the best we can do with a transferable model based upon the ESP method.

The average charge found for all monosaccharide carbon atoms is 0.228 with a standard deviation of 0.211. The standard deviation for the carbon atoms is unacceptably high, but when we consider each carbon atom separately we are able to significantly reduce this (see Table VII). For the polar and nonpolar hydrogens the values are 0.444 and 0.053 with standard deviations of 0.019 and 0.047, respectively. By considering each individual hydrogen atom we find that the standard deviation is not substantially improved; however, we do find a modest variation among the point charges. We have considered the methylene hydrogens as one pair because they are close in value and the average has a small standard deviation. Separating them has no significant affect on the results given below. Finally, the oxygen class had an average value of −0.660 with a standard deviation of 0.079. By considering each individual oxygen we find that the standard deviation is, in general, improved and that there is a significant variation among the point charges. The results of this analysis are presented in Table VII.

Table VII. Averages and standard deviations for 6-31G* (MNDO) charge values for a transferable charge model for monosaccharides.

Atom type	Average	SD
C(1-6)[a]	0.228 (0.136)	0.211 (0.105)
C(1)	0.477 (0.288)	0.168 (0.061)
C(2)	0.092 (0.086)	0.199 (0.068)
C(3)	0.274 (0.134)	0.114 (0.065)
C(4)	−0.036 (0.004)	0.115 (0.072)
C(5)	0.353 (0.147)	0.079 (0.072)
C(6)	0.206 (0.160)	0.051 (0.007)
H(20-24)[b]	0.444 (0.305)	0.019 (0.013)
H(20)	0.461 (0.321)	0.017 (0.009)
H(21)	0.435 (0.305)	0.020 (0.011)
H(22)	0.445 (0.307)	0.015 (0.010)
H(23)	0.437 (0.292)	0.021 (0.012)
H(24)	0.443 (0.299)	0.004 (0.004)
H(7-13)[c]	0.053 (0.054)	0.047 (0.033)
H(7)	0.056 (0.055)	0.028 (0.028)
H(8)	0.097 (0.073)	0.036 (0.019)
H(9)	0.036 (0.042)	0.042 (0.017)
H(10)	0.084 (0.070)	0.033 (0.016)
H(11)	0.079 (0.091)	0.040 (0.015)
H(12-13)	0.013 (0.016)	0.028 (0.021)
O(14-19)[d]	−0.660 (−0.453)	0.079 (0.051)
O(14)	−0.712 (−0.486)	0.025 (0.016)
O(15)	−0.671 (−0.471)	0.053 (0.030)
O(16)	−0.696 (−0.476)	0.043 (0.027)
O(17)	−0.639 (−0.436)	0.054 (0.027)
O(18)	−0.538 (−0.374)	0.094 (0.063)
O(19)	−0.703 (−0.476)	0.014 (0.007)

[a] All carbons.
[b] Polar hydrogens.
[c] Nonpolar hydrogens.
[d] All oxygens.

One can imagine several ways in which the point charges for a monosaccharide can be partitioned, but we have chosen to consider only two possibilities. From numerous tests we have found that the two approaches described here represent the best and worst cases for a chemically "sensible" transferable model. In the "coarse grain" model we considered only the C, H (polar and nonpolar), and O classes. In the second "fine grain" model we considered each individual charge separately (except for the methylene hydrogens). For each of these models we determined the dipole moment at the geometry used in the 6-31G* calculation for α and β-D-idose and compared it to the 6-31G* calculated dipole moment. To maintain the charge neu-

Table VIII. Dipole moments calculated from wave functions, ESP charges and transferable 6-31G* charges.

Method	μ_x	μ_y	μ_z	Total
α-D-Idose				
6-31G*	−2.81	−2.04	0.40	3.50
MNDO	−2.32	−1.73	0.49	2.94
6-31G* ESP	−2.80	−2.03	0.39	3.48
MNDO ESP	−1.81	−1.32	0.29	2.25
Scaled MNDO ESP (sugar only)[a]	−2.66	−1.93	0.42	3.31
Scaled MNDO ESP (all neutrals)[b]	−2.45	−1.79	0.39	3.05
Coarse model	−2.57	−4.92	2.34	6.02
Fine model	−3.08	−2.01	0.67	3.74
β-D-Idose				
6-31G*	2.76	−4.93	1.67	5.90
MNDO	2.22	−4.26	1.03	4.91
6-31G* ESP	2.76	−4.93	1.67	5.89
MNDO ESP	1.71	−3.36	0.95	3.85
Scaled MNDO ESP (sugar only)[a]	2.52	−4.95	1.40	5.73
Scaled MNDO ESP (all neutrals)[b]	2.32	−4.56	1.29	5.27
Coarse model	2.38	−8.75	2.34	9.36
Fine model	3.25	−5.97	0.94	6.86

[a]Scale factor of 1.47.
[b]Scale factor of 1.35.

trality of the monosaccharides using the transferable models we adopted the approach of Faerman and Price,[19] which is to at most vary the value of a charge by its standard deviation. In all cases studied we had to modify the charges only slightly. The results are given in Table VIII. In Table IX we calculate the rms deviation between the quantum mechanically determined electrostatic potential (6-31G*) and various classical models. These results reinforce the results of Table VIII. From these tables a few things are apparent: (1) Dipole moments calculated from 6-31G* ESP derived atomic point charges very closely match the calculated 6-31G* dipole moments; (2) dipole moments calculated by MNDO ESP derived point charges do not match the dipole moment calculated from the MNDO wave function; (3) dipole moments calculated from the MNDO ESP charges are lower than the 6-31G* dipole moments and scaled (by 1.47, for example) MNDO ESP charges reproduce the *ab initio* dipole moments reasonably well; (4) unscaled MNDO and the fine model reproduce the 6-31G* ESP potential relatively poorly, while scaled MNDO reproduces it nearly as well as the charges fitted to the 6-31G* ESP; (5) the coarse grain transferable model does very poorly; (6) the fine grain transferable model performs much better.

Table IX. Comparison of the 6-31G* ESP potential vs. several classical representations.

Comparison	Calculated rms deviation (kcal/mol)
α-D-Idose	
6-31G* vs. 6-31G* ESP	0.886
6-31G* vs. coarse transferable model	8.680
6-31G* vs. fine transferable model	3.060
6-31G* vs. MNDO ESP	4.731
6-31G* vs. scaled MNDO ESP (sugar only)[a]	1.393
6-31G* vs. scaled MNDO ESP (all neutrals)[b]	1.467
β-D-Idose	
6-31G* vs. 6-31G* ESP	0.888
6-31G* vs. coarse transferable model	9.824
6-31G* vs. fine transferable model	4.234
6-31G* vs. MNDO ESP	5.273
6-31G* vs. scaled MNDO ESP (sugar only)[a]	1.502
6-31G* vs. scaled MNDO ESP (all neutrals)[b]	1.838

[a]Scale factor of 1.47.
[b]Scale factor of 1.35.

Point (1) above indicates that 6-31G* ESP derived point charges do reproduce the 6-31G* calculated molecular dipole moment even though the charges were not forced to do so. This indicates that the ESP fitting procedure is capable of reproducing higher multipole moments (here the dipole) without forcing it to or resorting to multipole expansions. Point (2) indicates that MNDO does not reproduce the MNDO calculated dipole moment. This is due to the deorthogonalization of the semiempirical wave function.[7] Points (3) and (4) indicate that scaled MNDO performs quite well in reproducing 6-31G* dipole moments and electrostatic potentials, which suggests that MNDO in many cases should be the method of choice for the derivation of atomic point charges. We also have found that the MNDO electrostatic potentials match the 6-31G* ones very well and that the differences in them just involve a scale factor as is also true for the dipole moments. From points (5) and (6) we conclude that a coarse model is not appropriate to use in developing transferable point charges for monosaccharides. The fine grain model does much better and should be a useful approach in developing charge sets for monosaccharides. We have looked into improving this transferable model by scaling the charges. However the scale factor is 0.999, which mitigates the usefulness of the scaling procedure in this case.

The preceding analysis clearly shows that "raw" MNDO ESP charges and the fine transferable charge model are of about equal reliability. Scaling the MNDO ESP charges results in a charge model that is capable of nearly reproducing the quality of the 6-31G* ESP derived point charges. This leads to the conclusion that scaled MNDO ESP charges are a powerful way in which to rapidly derive atomic point charges of a reasonable quality for routine uses. The magnitude we are seeing in the

Table X. Averages and standard deviations of the ESP charges derived for main-chain atoms.

Atom number and type	Average	SD
C5 (Cα)	−0.149	0.244
C2&6	0.756	0.100
C2	0.821	0.050
C6	0.691	0.095
C1	−0.501	0.034
C9	−0.269	0.029
O3&7	−0.628	0.042
O3	−0.661	0.026
O7	−0.595	0.024
N4&8	−0.491	0.088
N4	−0.522	0.108
N8	−0.460	0.048
H10-12	0.135	0.014
H14	0.128	0.056
H16-18	0.136	0.012
H13-15	0.332	0.031

calculated rms deviation in reproducing an electrostatic potential using our transferable model (3-4 kcal/mol) is much higher than that observed by, for example, Faerman and Price using a distributed multipole analysis (DMA).[19] This suggests that the DMA approach might be more robust in defining an accurate transferable charge model. However, these authors only examined a small data base, which might not have been able to fully assess that abilities of the DMA approach. Finally, the monosaccharides provided a straightforward way to develop a transferable charge model, but the amino acids are more difficult and also illustrate more problems associated with using the ESP technique in the development of an atom-by-atom transferable model.

Amino Acids

The amino acids present a much more complicated problem when trying to develop a transferable charge model because of the numerous functional groups present in the side chains. The amino acids also pose further problems to developing an ESP based transferable model. To demonstrate some of these problems we will first consider the constant backbone portion of the amino acids.

Table X gives the 6-31G* ESP charges for the main-chain atoms of our dipeptide models for the various naturally occurring amino acids. All of the values have relatively low standard deviations except for Cα, which has a standard deviation of 0.244. The large variation of atomic charges for Cα has also been observed when considering several conformations of alanine and was postulated to be due to variations in nonbonded contacts between the conformers.[29] We also find that this is dependent upon the side-chain characteristics (the backbone conformations we consider are the same in all cases). This phenomenon is best described by considering glutamic and aspartic acid. In Glu there are two methylenes between the carboxyl carbon and Cα, while in Asp there is only one. The carboxyl carbon is positive and the next attached methylene is negative, the next positive, and so on. The result is that in Glu Cα is negative (−0.419), while in Asp it is positive (0.085). Consideration of other side chains gives a similar picture (see Table XI). Thus, Cα is a position where a large variability is present due to both conformational effects[29] and to the character of the side chain. By examining Table XI we find it is hard to come up with a realistic way in which to categorize the charge on Cα with the character of the side chain (i.e., aromatic charged, etc.). For example, consider the alkyl sidechains of Ala, Val, Leu, and Ile, which might be expected to have similar Cα charges. These vary from 0.228 for Ala to −0.496 for Val, a range of over 0.7 charge units. If we eliminate Ala the range is still 0.3 units of charge. Thus, we conclude that we should either use the average value given in Table X or use each individual value for Cα.

Table XI. Variation in the Cα charge with amino acid side chain.

Residue	Cα Charge
Ala	0.22788
Asn	−0.01159
Asp	0.08479
Cys	−0.17434
Gln	−0.59121
Gly	−0.43787
Arg	−0.02447
Hie	−0.02600
Hid	0.19555
Hip	−0.03988
Ile	−0.39632
Leu	−0.19742
Lys	−0.12120
Met	−0.00969
Pro	−0.30071
Phe	0.29245
Ser	−0.27091
Trp	−0.26095
Tyr	0.03714
Val	−0.49603
Glu	−0.41858
Thr	−0.33104

While Cα has the largest standard deviation in Table X it is interesting to note that groups attached to Cα have the next largest set of standard deviations. For example, C6 flanks Cα and has a much larger standard deviation than does C2 which is in a similar chemical environment. The same is true for N4 vs. N8 and H14 has the largest standard deviation of all hydrogens. Thus, the large variation at Cα affects groups attached to it quite significantly.

The side-chain groups pose a major difficulty in

developing a transferable model because in many instances we have only one example for an atom type. This makes it hard to get statistically meaningful average charges. To improve this situation many small representative molecules can be considered,[19] but our present data base is not complete enough to address this issue. Given that our *main-chain* charges are the only statistically meaningful results we have, we feel that the only logical way to get a transferable model based upon our average main-chain charges is to evaluate ESP charges for the side chain only using 6-31G* or MNDO. These can then be pieced together with the main-chain charges using the approach described by Faerman and Price.[19] This is what was essentially done in the original AMBER force field.[48] Finally, we feel that all the difficulties we have encountered with the monosaccharides and amino acids make it unlikely that a transferable model based upon the ESP approach can be developed beyond the accuracy indicated here. Given the importance of the amino acid residues we have considered it is best to try to develop a transferable charge model based upon the residue concept described at the beginning of this section. This approach will be described shortly and will be employed in a second generation AMBER force field.[31,44]

With the amino acids we encounter charged species. It has not been assessed how well MNDO does with charged molecules previously, so comments are warranted here.[7] We find that MNDO does quite well for charged species. Considering charged amino acids alone we determine a correlation coefficient of 0.92 and a scale of only 1.08. This is fortunate because scaling a charged species results in a net total charged increased by the multiplicative scale factor. From this preliminary observation it appears best to use MNDO charges directly for charged molecules. However, more testing should be done before a definitive conclusion can be reached.[45]

Scaled MNDO performed so well with the monosaccharides we might expect the same for the neutral amino acid residues. Recall, however, that nitrogen was a problem for MNDO, so it is likely MNDO will not perform as well as it did for the monosaccharides. Table XII gives the dipole moment information while Table XIII gives the rms fit of the various point charge models to the 6-31G* electrostatic potential of tryptophan. We indeed see that for the amino acids MNDO and scaled MNDO do not perform as well as they did for the monosaccharides. Furthermore we find that the scaling of the charges does not dramatically improve the results. This is likely due to the problems with nitrogen. One might imagine scaling the nitrogen atoms separately, but this leads to a nonintegral charge and is, therefore, not the best approach to adopt when developing a point charge model to be used in molecular simulations.

Table XII. Dipole moments calculated from wave functions and ESP charges.

Method	Trp			
	μ_x	μ_y	μ_z	Total
6-31G*	0.90	3.55	1.29	3.88
MNDO	0.39	3.09	1.40	3.42
6-31G* ESP	0.88	3.59	1.28	3.91
MNDO ESP	0.42	2.88	1.32	3.20
Scaled MNDO ESP (amino acid only)[a]	0.53	3.63	1.66	4.03
Scaled MNDO ESP (all neutrals)[b]	0.57	3.89	1.78	4.32

[a]Scale factor of 1.26.
[b]Scale factor of 1.35.

Fluctuations in ESP Derived Point Charges

Figure 3 gives a Φ, Ψ plot of the conformational space that was sampled during the course of our QMD simulation. We find that the α-helical region is sampled a bit more than the β-sheet region, by about 10 ps. Figure 4 further illustrates this. This figure gives the fluctuation of the dipole moment of the alanine dipeptide during the course of the MD simulation. The α-helical structure has a dipole moment of $\sim$5 D, while the β-sheet conformation has a dipole moment of $\sim$2.0 D. Figure 4 clearly shows that up to about 30 ps we have an α-helical structure and between 30 ps to $\sim$33 ps there is a transition to the dipole moment that characterizes the β-sheet. The β-sheet conformation then persists throughout the remainder of the simulation. By separately mapping out the Ramachandran plot for the dipeptide we found that MNDO favors the β-sheet conformation over the α-helical structure by about 5 kcal/mol. Thus, it is not surprising that after a period of time the QMD simulation carries us into the β-sheet region. Table XIV gives the average charges and the associated standard deviations (σ) for the whole simulation (50 ps) and the first 25 ps (α-helical region) and the last 25 ps of

Table XIII. Comparison of the 6-31G* ESP potential vs. several classical representations.

Comparison	Trp calculated rms deviation (kcal/mol)
6-31G* vs. 6-31G* ESP	0.915
6-31G* vs. MNDO ESP	2.798
6-31G* vs. scaled MNDO ESP (amino acid only)[a]	2.288
6-31G* vs. scaled MNDO ESP (all neutrals)[b]	2.741

[a]Scale factor of 1.26.
[b]Scale factor of 1.35.

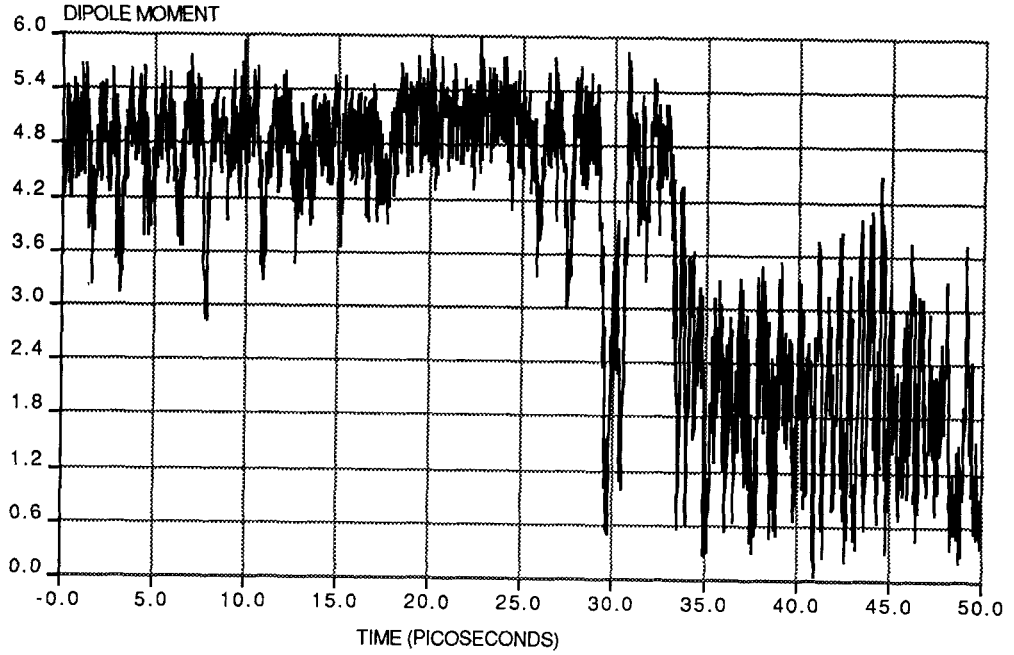


Figure 3. Ramachandran plot of the values of Φ and Ψ sampled during the QMD simulation.

the simulation (predominantly β-sheet). Table XIV also gives the averages for the molecular dipole moment.

As can be seen from the table the *average* computed charges for the three sets only vary by about ±0.01 e, while the dipole moments vary by a relatively large amount. The dipole moments vary widely because we are changing the molecular conformation from α-helical to the more extended β-sheet conformations (see Fig. 4). Interestingly, though, the charges do not vary significantly from one rotational conformer class (α → β) to another.

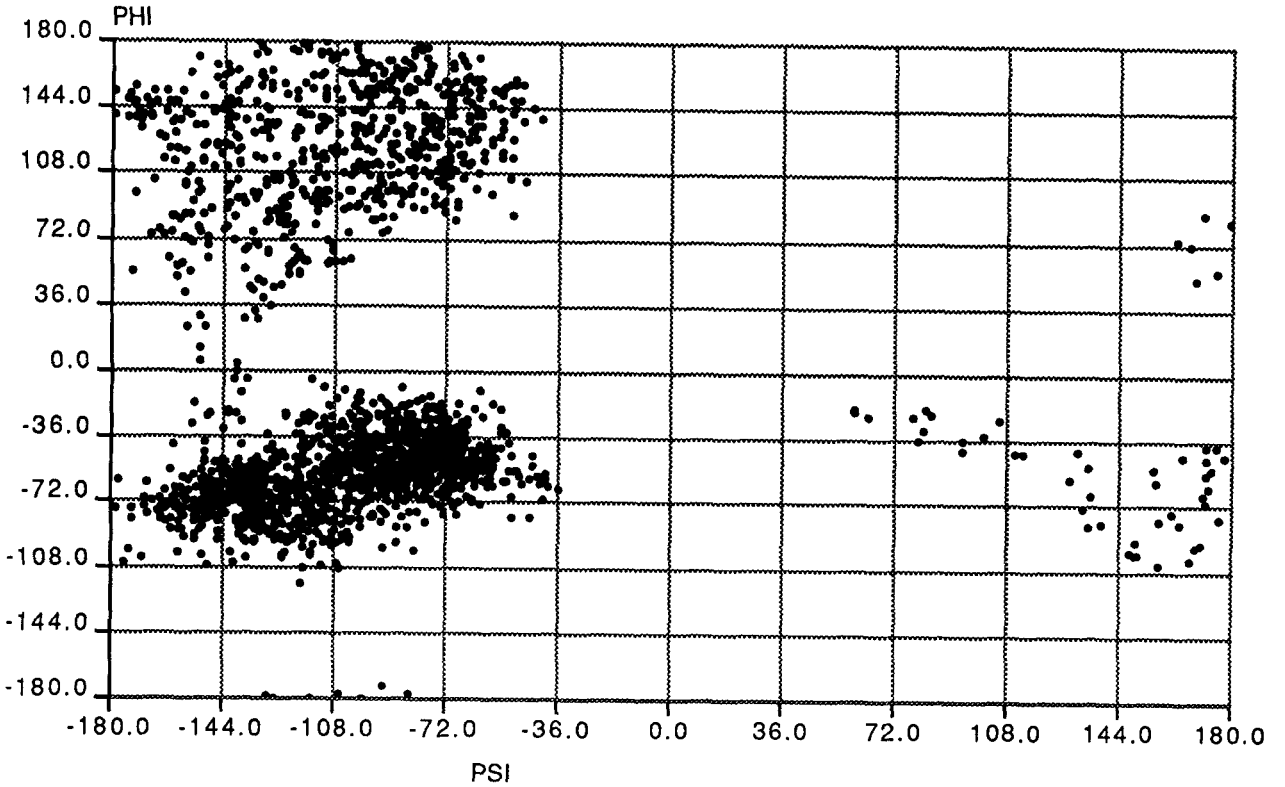


Figure 4. Fluctuation of the calculated dipole moment during the course of the QMD simulation.

Table XIV. Variation of ESP calculated point charges during the QMD simulation.

	Ala					
Atom label	50 ps (AVE)	50 ps (σ)	25 ps[a] (AVE)	25 ps (σ)	25 ps[b] (AVE)	25 ps (σ)
C−1	−0.3646	0.0471	−0.3681	0.0478	−0.3611	0.0463
C−2	0.6723	0.0500	0.6702	0.0496	0.6744	0.0502
O−3	−0.4421	0.0176	−0.4369	0.0155	−0.4474	0.0181
N−4	−0.6843	0.0770	−0.6751	0.0780	−0.6934	0.0750
C−5	0.1449	0.1080	0.1281	0.1120	0.1617	0.1010
C−6	0.5407	0.0683	0.5393	0.0685	0.5421	0.0680
O−7	−0.4197	0.0178	−0.4217	0.0180	−0.4177	0.0174
N−8	−0.5574	0.0662	−0.5322	0.0559	−0.5826	0.0662
C−9	0.0257	0.0859	−0.0019	0.0712	0.0533	0.0903
H−10	0.1051	0.0164	0.1049	0.0168	0.1054	0.0160
H−11	0.1062	0.0167	0.1074	0.0166	0.1050	0.0167
H−12	0.1053	0.0171	0.1057	0.0176	0.1050	0.0166
H−13	0.2958	0.0284	0.2899	0.0284	0.3016	0.0272
H−14	0.0719	0.0302	0.0784	0.0295	0.0654	0.0294
H−15	0.2772	0.0225	0.2724	0.0224	0.2821	0.0215
H−16	0.0444	0.0289	0.0516	0.0273	0.0372	0.0287
H−17	0.0464	0.0282	0.0499	0.0269	0.0429	0.0291
H−18	0.0489	0.0300	0.0551	0.0279	0.0428	0.0307
C−19	−0.2172	0.0777	−0.2132	0.0828	−0.2211	0.0721
H−20	0.0618	0.0288	0.0574	0.0289	0.0662	0.0281
H−21	0.0726	0.0239	0.0779	0.0211	0.0672	0.0252
H−22	0.0659	0.0263	0.0609	0.0270	0.0710	0.0245
μ (debye)	3.81	1.52	4.87	0.47	2.75	1.46

[a]First 25 ps of sampling.
[b]Second 25 ps of sampling.

This leads one to believe that determining the charges in one molecular conformation should be reasonably transferable to another that may arise during the course of a MD simulation using a classical force field. However, as has been shown by a number of authors this is not correct.[19,22,29,30] Another interesting feature to note is that the variability of the three average atomic charge sets in Table XIV is less than that of the standard deviation computed for each set. This just indicates that the fluctuations in the charges can be quite significant, but that the average values of the charges converges very quickly. All of this data taken together raises the question of where does all of the charge variability come from given that it is not clearly a conformational effect. First one might expect this to come from an orientational dependence in the ESP calculations; however, we have shown that this is not a big factor for our calculations (see Table I). It is possible that this difference is due to an admixture of the α and β conformers in the last 25 ps of the simulation. To eliminate this we considered only the data sets that were clearly generated by a β-sheet conformation (last 15 ps). We find that the variation is similar to that given in Table XIV. Thus, the fluctuations are not due to poor statistical averaging of the α and β parts of the simulation.

Given that the charge variation is not solely caused by conformational effects nor is it caused by an orientation dependence one is left to conclude that the charge variability comes from local changes during the simulations—these being changes in bond lengths and angles that naturally occur during the course of a simulation. Another important source of charge variation likely comes from induced polarization effects. All of these could be more important than (or as important as) the conformational effects and therefore it is possible that the latter is being masked by these other effects. We attempted to correlate bond and angle fluctuations with charge fluctuations, but found only poor correlations. This is consistent with what we found for the conformational analysis. Taken as a whole this data suggests that many effects (bond, angle, and torsional fluctuations as well as polarization effects) are operative and that it is not possible to identify any one as the dominant reason for the charge variation in the present case.

From these simulations we can estimate how these effects would influence the structure and dynamics of the alanine dipeptide during the course of an MD simulation. For example, we can estimate the magnitude of these effects for hydrogen bond formation. We evaluate this by considering the electrostatic portion of a classical force field [E_{elec} (kcal/mol) = $332q_iq_j/\varepsilon R_{ij}$] only and determining the maximum and minimum values that would be seen for this term given the average charges and the σ values for these charges (Table

XIV). Taking C=O (atom numbers 2 and 3) and N−H (8 and 15) charges from Table XIV and assuming a constant dielectric of one, a H · · · O distance of 2.0 Å, a N−H distance of 1:0 Å, a C=O distance of 1.32 Å, and correcting for orientational effects (see Table I) we arrive at a maximum value of −4.0 kcal/mol, a minimum of −2.6 kcal/mol, and an average electrostatic interaction energy of −3.2 kcal/mol. This is a rather large range of values and indicates that the strength of a hydrogen bond can fluctuate quite significantly during the course of a simulation without the distance between the donor and acceptor atoms changing. In this analysis we have, of course, neglected other terms in the force field that would affect hydrogen bond strength (i.e., Lennard-Jones terms).

From this simulation it is clear that the determination of ESP derived atomic point charges at a fixed geometry will not give you an accurate description of how the electron density will be distributed during the course of a molecular dynamics simulation. To partially alleviate this problem average charges obtained using the QMD approach described here or via averaging over an ensemble of conformations might give a better representation of the electron distribution in a conformationally flexible molecule. This would give a better average picture of the electron distribution. Inclusion of terms into force fields that model polarization effects* would also alleviate this problem, but lead to a significant increase in the expense of any given calculation. Finally, the results of this simulation or others like it could be used to study how well a polarizable model reflects how electron density is redistributed during the course of an MD simulation.

CONCLUSIONS

We have shown that our approach to determining ESP derived atomic point charges using *ab initio* or semiempirical techniques does not suffer from a dependence upon the molecular orientation in 3D space and we have shown that our standard approach is "converged" relative to cases where we use a higher density of points. Second, we clearly demonstrated that ESP charges derived using the MNDO deorthogonalized wave function are a reasonable approximation to ESP point charges determined at the 6-31G* level. Furthermore scaling of these charges can in many cases drastically improve the agreement. We have reported a number of possible scale factors, but we recommend that a value of 1.35 be used for cases in which C, H, N, and O atoms are present in the molecule under consideration. For a molecule with only C, H, and O a value of 1.47 is recommended. For a charged molecule we recommend that the charges be left unscaled, but this needs to be further studied.[45] Finally, even given the reasonableness of the MNDO ESP approach we still recommend the use of high quality *ab initio* ESP derived atomic point charges whenever high accuracy is needed.

It should be stressed that 6-31G* electrostatic derived atomic point charges are not definitive,[51] but that they have several features that make them attractive for use in computer simulations using effective two-body potentials. From extensive work on the development of effective two-body water models it has been found that the dipole moment of water should be approximately 10% above the gas-phase dipole moment.[52] This is thought to be necessary to mimic polarization effects that are important when simulating liquid water.[7,49,50] ESP derived atomic charges at the 6-31G* level satisfy this criteria[4] and as such make this level of approximation attractive for people carrying out molecular simulations.[7] Preliminary evidence from this lab indicates that ESP derived atomic point charges do not change much as we increase the basis set size.[53] Furthermore, we do not see much variation as we include correlation at both the CI and Møller-Plasset level.[53]

We developed an ESP based transferable charge model for monosaccharides based upon an atom-by-atom consideration, which is similar to the decomposition used by others.[19,46,47] This procedure performs quite well, but MNDO performs almost as well and it appears to be superior when scaled. Unfortunately, scaling does not improve the transferable charge model. Given the fluctuations in charges observed in the QMD simulation and that observed by others using the ESP approach[22,29,30] it seems reasonable to suggest that MNDO is as useful as a transferable approach. This is further warranted by the fact that MNDO derived ESP charges can be determined relatively quickly and can be used to evaluate charges for large molecules (200+ atoms). However, for very large molecules (i.e., proteins) a transferable model is necessary be it an atom-by-atom model, functional group model, or a residue based approach and the use of high quality *ab initio* ESP calculations to derive the transferable model makes the most sense.

We carried out a QMD simulation that shows that rather substantial charge variation can be expected during an MD simulation. This is not a surprising result and has been described by others.[19,22,29,30] Nevertheless, it is a bit unsettling since most simulations assume a fixed charge distribution. How large of an affect this charge variation might have on the dynamics of a large protein

*For treatments of intermolecular polarization see ref. 49 and references cited therein. For treatments of intramolecular polarization effects see ref. 50.

is currently unknown, but implicit inclusion of polarization effects in an MD simulation of a protein and comparing this to one without would be instructive. The QMD method employed the MNDO method, which as we have indicated here is capable of reasonably mimicking *ab initio* electrostatic potentials. It would be preferable to use *ab initio* techniques in this evaluation, but given the great expense of these methods and the ability of MNDO to reproduce high quality *ab initio* electrostatic potentials it would appear that we are justified in using MNDO to study intramolecular polarization effects.

The body of evidence presented here (and elsewhere[19,22,29,30]) suggests that the determination of atomic point charges using the ESP technique results in a set of charges that can vary by a significant amount depending upon the molecular geometry considered. This variation depends upon a number of factors such as fluctuations in bonds and angles, changes in molecular conformation, and on induced polarization effects. Incorporation of these effects into classical potentials appears necessary to greatly improve the coulombic electrostatic representation commonly employed in current generation force fields.*

The San Diego Supercomputer Center is acknowledged for supplying significant amounts of computer time and the NSF is acknowledged for supplying funds to purchase two Multiflow 14/300 superminicomputers. Helpful discussions with Charlie Brooks of Carnegie Mellon University are also acknowledged.

References

1. S.K. Burley and G.A. Petsko, *Adv. Protein Chem.*, **39**, 125 (1988).
2. M. Davis and J.A. McCammon, *Chem. Rev.*, **90**, 509 (1990).
3. S.C. Harvey, *Proteins: Struct. Func. Genet.*, **5**, 78 (1989).
4. W.J. Hehre, L. Radom, P.v.R. Schleyer, and J.A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York, 1986.
5. S.R. Cox and D.E. Williams, *J. Comp. Chem.*, **2**, 304 (1981).
6. D.E. Williams and J.M. Yan, *Adv. Atomic Mol. Phys.*, **23**, 87 (1988).
7. B.H. Besler, K.M. Merz Jr., and P.A. Kollman, *J. Comp. Chem.*, **11**, 431 (1990).
8. P.L. Cummins and J.E. Gready, *Chem. Phys. Lett.*, **174**, 355 (1990).
9. G.G. Freenczy, C.A. Reynolds, and W.G. Richards, *J. Comp. Chem.*, **11**, 159 (1990).
10. J.M. Gruschus and A. Kuki, *J. Comp. Chem.*, **11**, 978 (1990).
11. M. Orozco and F.J. Luque, *J. Comp. Chem.*, **11**, 909 (1990).
12. W.L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).
13. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.*, **4**, 187 (1983).
14. P. Coppens, *J. Phys. Chem.*, **93**, 7979 (1989).
15. D.A. Pearlman and S.-H. Kim, *Biopolymers*, **24**, 327 (1985).
16. D.A. Pearlman and S.-H. Kim, *J. Mol. Biol.*, **211**, 171 (1990).
17. S.L. Price, R.J. Harrison, and M.F. Guest, *J. Comp. Chem.*, **10**, 552 (1989).
18. A.J. Stone and M. Alderton, *Mol. Phys.*, **56**, 1047 (1985).
19. C.H. Faerman and S.L. Price, *J. Am. Chem. Soc.*, **112**, 4915 (1990).
20. M.J.S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899 (1977).
21. M.J.S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4907 (1977).
22. C.M. Breneman and K.B. Wiberg, *J. Comp. Chem.*, **11**, 361 (1990).
23. M.L. Connolly, *J. Appl. Cryst.*, **16**, 548 (1983).
24. L.E. Chirlian and M.M. Francl, *J. Comp. Chem.*, **8**, 894 (1987).
25. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
26. J.J.P. Stewart, *J. Comp. Chem.*, **10**, 209 (1989).
27. J.J.P. Stewart, *J. Comp. Chem.*, **10**, 221 (1989).
28. J.J.P. Stewart, *J. Comp. Chem.*, **12**, 320 (1991).
29. D.E. Williams, *Biopolymers*, **29**, 1367 (1990).
30. K.B. Wiberg, C.M. Hadad, C.M. Breneman, K.E. Laidig, M.A. Murcko, and T.J. LePage, *Science (Wash. DC)*, **252**, 1266 (1991).
31. K.M. Merz Jr., unpublished results.
32. M.J. Frisch, M. Head-Gordon, H.B. Schlegel, K. Raghavachari, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R. Martin, L.R. Kahn, J.J.P. Stewart, E.M. Fluder, S. Topiol, and J.A. Pople, *Gaussian 86*, Gaussian Inc., Pittsburgh, PA, 1988.
33. U.C. Singh and P.A. Kollman, *J. Comp. Chem.*, **5**, 129 (1984).
34. J. Almlöf, J.K. Faegri, and K. Korsell, *J. Comp. Chem.*, **3**, 385 (1982).
35. F.H. Allen, S. Belard, M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B.G. Hummelink-Peters, O. Kennard, W.D.S. Motherwell, J.R. Rodgers, and D.G. Watson, *Acta. Crystallogr.*, **B35**, 2331 (1979).
36. F.H. Allen, O. Kennard, D.G. Watson, L. Brammer, A.G. Orpen, and R. Taylor, *J. Chem. Soc., Perkins II*, S1 (1987).
37. G.M. Brown and H.A. Levy, *Science*, **147**, 1038 (1965).
38. K.M. Merz Jr. and B.H. Besler, *QCPE Bull.*, **10**, 15 (1990).
39. R. Car and M. Parrinello, *Phys. Rev. Lett.*, **55**, 2471 (1985).
40. R.O. Jones and D. Hohl, *J. Am. Chem. Soc.*, **112**, 2590 (1990).
41. J.J.P. Stewart, L.P. Davis, and L.W. Burggraf, *J. Comp. Chem.*, **8**, 1117 (1987).
42. P. Bash, M. Fields, and M. Karplus, *J. Am. Chem. Soc.*, **109**, 8092 (1987).
43. M.J.S. Dewar and Y. Yamaguchi, *Comp. Chem.*, **2**, 25 (1978).

*Supplementary material is available from the author upon request. This includes coordinates (*XYZ*) for the various molecules studied as well as all of the computed ESP atomic point charges.

44. P.A. Kollman and K.M. Merz Jr., manuscript in preparation.
45. T.M. Glennon and K.M. Merz Jr., research in progress.
46. R.F.W. Bader, A. Larouche, C. Gatti, M.T. Carroll, P.J. MacDougall, and K.B. Wiberg, *J. Chem. Phys.*, **87,** 1142 (1987).
47. M. Eisenstein, *Int. J. Quantum Chem.*, **33,** 127 (1988).
48. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, *J. Am. Chem. Soc.*, **106,** 765 (1984).
49. M. Sprik and M.L. Klein, *J. Chem. Phys.*, **89,** 7556 (1988); J. Caldwell, L.X. Dang, and P.A. Kollman, *J. Am. Chem. Soc.*, **112,** 9144 (1990).
50. U. Dinur, *J. Phys. Chem.*, **94,** 5669 (1990).
51. F.J. Luque, M. Orozco, F. Illas, and J. Rubio, *J. Am. Chem. Soc.*, **113,** 5203 (1991).
52. W.L. Jorgensen, J. Chandrasekhar, J. Madura, R.W. Impey, and M.L. Klein, *J. Chem. Phys.*, **79,** 926 (1983).
53. T.M. Glennon and K.M. Merz Jr., research in progress.