Accurate Modeling of the Intramolecular Electrostatic Energy of Proteins

MICHAEL J. DUDEK and JAY W. PONDER*
Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received 25 March 1994; accepted 22 August 1994

ABSTRACT

The $(\phi, \psi)$ energy surface of blocked alanine ($N$-acetyl-$N'$-methyl alanineamide) was calculated at the Hartree-Fock (HF)/6-31G* level using *ab initio* molecular orbital theory. A collection of six electrostatic models was constructed, and the term *electrostatic model* was used to refer to (1) a set of atomic charge densities, each unable to deform with conformation; and (2) a rule for estimating the electrostatic interaction energy between a pair of atomic charge densities. In addition to two partial charge and three multipole electrostatic models, this collection includes one extremely detailed model, which we refer to as nonspherical CPK. For each of these six electrostatic models, parameters—in the form of partial charges, atomic multipoles, or generalized atomic densities—were calculated from the HF/6-31G* wave functions whose energies define the *ab initio* energy surface. This calculation of parameters was complicated by a problem that was found to originate from the locking in of a set of atomic charge densities, each of which contains a small polarization-induced deformation from its idealized unpolarized state. It was observed that the collective contribution of these small polarization-induced deformations to electrostatic energy differences between conformations can become large relative to *ab initio* energy differences between conformations. For each of the six electrostatic models, this contribution was reduced by an averaging of atomic charge densities (or electrostatic energy surfaces) over a large collection of conformations. The *ab initio* energy surface was used as a target with respect to which relative accuracies were determined for the six electrostatic models. A collection of 42 more complete molecular mechanics models was created by combining each of our six electrostatic models with a collection of seven models of repulsion + dispersion + intrinsic torsional energy, chosen to provide a representative sample of functional forms and parameter sets. A measure of distance was defined between model and *ab initio* energy surfaces; and distances were calculated for each of our 42 molecular mechanics models. For most of our 12 standard molecular mechanics models, the average error between model and *ab initio* energy surfaces is greater than 1.5 kcal/mol. This error is decreased by (1) careful treatment of the nonspherical nature of atomic charge densities, and (2)
accurate representation of electrostatic interaction energies of types 1—2 and 1—3. This result suggests an electrostatic origin for at least part of the error between standard model and \textit{ab initio} energy surfaces. Given the range of functional forms that is used by the current generation of protein potential functions, these errors cannot be corrected by compensating for errors in other energy components. © 1995 by John Wiley & Sons, Inc.

\textbf{Introduction}

I\n this article we look at the effect on the \((\phi, \psi)\) energy surface of blocked alanine of substituting increasingly accurate representations of the electrostatic energy. More specifically, we (1) calculate a target \textit{ab initio} energy surface; (2) gather a collection of six electrostatic models, differing with respect to the approximations that are introduced and ranging in accuracy from crude to nearly exact; (3) combine these six electrostatic models with seven models of repulsion + dispersion + intrinsic torsional energy to create a collection of 42 complete molecular mechanics models; (4) define a measure of distance between two energy surfaces; and finally (5) look at the distance between model and \textit{ab initio} energy surfaces as a function of electrostatic model. We show that, over a small collection of widely varying repulsion models, the distance between model and \textit{ab initio} energy surfaces decreases as the representation of short-range electrostatic interaction energies becomes more accurate. This result suggests the existence of an energy contribution that has not yet been accounted for in any of the current generation of protein potential functions.\textsuperscript{1–7}

For small nonpolar organic molecules, the best current molecular mechanics force fields allow calculation of structure and energetics to within experimental accuracy.\textsuperscript{7,34,35} For protein molecules, which contain both polar functional groups and adjacent torsion angle degrees of freedom, the accuracy of existing molecular mechanics force fields relative to that needed to enable structure prediction is not easily evaluated. The problem of obtaining a molecular mechanics-based method for predicting protein structure can be divided into two primary subproblems: the multiple-minima problem, and the problem of obtaining an accurate representation of the energy surface of a protein and its solvent environment. For some of the handful of cases in which the multiple-minima problem has been solved, predicted and experimentally observed structures have failed to agree, suggesting error in the energy functions that were used.\textsuperscript{9–13}

Analysis of errors in predicted structures has suggested that the goal of reliable structure prediction will not be obtained without the introduction of more accurate representations for both the vacuum potential energy and the hydration free energy.\textsuperscript{9} In this article, we focus on the vacuum potential energy. Within the framework of molecular mechanics,\textsuperscript{14} components of the vacuum potential energy include electrostatic, polarization, repulsion, intrinsic torsional, and dispersion energies, as well as energies resulting from bond stretching, bond angle bending, and possible coupling between stretching and bending. Each of these components is a possible source of error. In this article we focus on the electrostatic energy. More specifically, we focus on short-range intramolecular electrostatic interaction energies between atomic charge densities that deviate from spherical symmetry, a component that has little or no significance in molecules that contain no torsion angle degrees of freedom.

From quantum mechanics, we know that the charge density of a molecule is not a superposition of many spherically symmetric atomic charge densities.\textsuperscript{15,16} In particular, quantum mechanics predicts a buildup of electron density at the center of each covalent bond relative to the density that results from superposition of spherically symmetric atomic charge densities. The electrostatic interaction energy between generalized atomic charge densities can be represented exactly using a multipole expansion, provided that the distance between nuclei is greater than the sum of the radii of the atomic densities.\textsuperscript{17} A method for obtaining atomic multipoles from wave functions has been suggested by Stone.\textsuperscript{18} For small-molecule dimers, Buckingham and Fowler have shown that reliable prediction of experimentally observed geometries requires an accurate (multipole expansion truncated at quadrupole) representation of intramolecular electrostatic interaction energies.\textsuperscript{19,20}
Atom pairs that are separated by one, two, or three bonds will be referred to as 1—2, 1—3, or 1—4, respectively. For electrostatic interaction energies of types 1—2 and 1—3, representations that account for the nonspherical nature of atomic charge densities predict a dependence on the torsion angles that determine the relative orientation of the interacting atomic densities. Because the internuclear distances are too small, a multipole-expansion representation of this dependence is not accurate. Also, as a consequence of smaller internuclear distances, the effects of anisotropic atomic charge densities on energy differences between conformations are expected to be larger for 1—2 and 1—3 electrostatic interaction energies than for intermolecular electrostatic interaction energies.

**Methods**

**THEORETICAL FRAMEWORK**

We introduce the notation \([q_m]\) to represent energy calculated using ab initio molecular orbital theory as a function of nuclear coordinates. For a molecule, natural components of the quantum mechanical energy include kinetic; nuclear—nuclear repulsion; electron—nuclear attraction; and electron—electron repulsion, which can be further decomposed into Coulomb and exchange integrals.\(^{21}\) We have chosen to combine these basic components to obtain the less fundamental decomposition

\[
[q_m] = [\text{kin}] + [\text{coul}] + [\text{exch}]
\]  

(1)

Here \([\text{kin}]\) is defined to include the kinetic energy integrals, and \([\text{coul}]\) is defined to include nuclear—nuclear repulsion; the electron—nuclear attraction integrals; and the electron—electron repulsion Coulomb integrals, including

\[
\frac{1}{2} \sum_{j=1, \ldots, N} \int \psi_j^*(x_1) \psi_j(x_1) \frac{1}{|r_2 - r_1|} \times \psi_j^*(x_2) \psi_j(x_2) \, dx_1 \, dx_2
\]  

(2)

where \(\psi_j\) is the jth molecular spin orbital, \(x_1\) the space-spin variable of electron 1, \(r_1\) the space component of \(x_1\), and \(N\) the number of electrons. Moreover, \([\text{exch}]\) is defined to include the electron—electron repulsion exchange integrals, including the negative of eq. (2). The decision to include eq. (2) in the definition of \([\text{coul}]\) causes \([\text{coul}]\) to be equal to the classical electrostatic energy of the system consisting of the nuclei and the electron density described by the wave function. This gives us a natural decomposition into the electrostatic energy of a classical charge density \([\text{coul}]\) and quantum effects \([\text{kin}] + [\text{exch}]\).

In a molecular mechanics treatment, natural components of the energy include electrostatic, polarization, repulsion, intrinsic torsional, and dispersion energies, in addition to the energy that is required to deform bond lengths and bond angles from their low-energy positions.\(^{14}\) Again, we find it convenient to combine these basic components to obtain the less fundamental decomposition

\[
[q_m] = [\text{elec}] + [\text{pol}] + [\text{rep}] + [\text{tor}] + [\text{geom}]
\]  

(3)

where \([\text{elec}]\) is the electrostatic energy, \([\text{pol}]\) the polarization energy, \([\text{rep}]\) the sum of repulsion and dispersion energies, \([\text{tor}]\) the intrinsic torsional energy, and \([\text{geom}]\) the energy required to deform bond lengths and bond angles. In this article, for each of the systems that we look at, the domain over which \([q_m]\) is calculated is restricted to a set of conformations for which all bond lengths and bond angles are maintained rigid. Because the molecular mechanics component \([\text{geom}]\) is constant over these sets of conformations, it will not be considered further. An isolated change to one of the four remaining molecular mechanics components, caused by a change in conformation, is expected to correspond to a characteristic pattern of change in the quantum mechanics components. In Table I, we introduce notation for quantum mechanics subcomponents of molecular mechanics components and vice versa.

Next, we will use the wave functions and the \([\text{coul}]\) component of an ab initio energy surface to construct reasonable estimates of \([\text{elec}]\), \([\text{pol}]\), and \([\text{rep}]\)—the \([\text{coul}]\) subcomponents of \([\text{elec}]\), \([\text{pol}]\), and \([\text{rep}]\) + \([\text{tor}]\), respectively. This decomposes \([\text{coul}]\), the quantum mechanics component corresponding to the classical electrostatic energy of the charge density, into molecular mechanics subcomponents.

**Table I.** Notation for Quantum Mechanics Subcomponents of Molecular Mechanics Components and Vice Versa.

<table>
<thead>
<tr>
<th>Molecular Mechanics Components</th>
<th>Quantum Mechanics Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>([\text{elec}])</td>
<td>([\text{pol}])</td>
</tr>
<tr>
<td>([\text{pol}])</td>
<td>([\text{elec}])</td>
</tr>
<tr>
<td>([\text{rep}] + [\text{tor}])</td>
<td>([\text{tor}])</td>
</tr>
</tbody>
</table>

\([\text{kin}]\)  —    \([\text{tor}]\)
\([\text{exch}]\) —    \([\text{exch}]\)
\([\text{coul}]\) \([\text{elec}]\) \([\text{pol}]\) \([\text{tor}]\)
Later in the Methods section, these constructions will be useful for clarifying the distinction between polarization energy, \([ pol \]), and electrostatic stabilization caused by the use of atomic charge densities that contain polarization-induced deformations from their idealized unpolarized states.

We assume that \([ elec \]), the electrostatic component of an energy surface, is obtained by proper translation and rotation of an idealized set of unpolarized atomic charge densities, each atomic density being translated and rotated along with the fragment defined by its nucleus and the set of connected nuclei. Polarization or other flow of electron density with conformation does not contribute. To obtain an estimate of \([ elec \]), we first construct a collection of energy surfaces \([ coul \]), in a sense extending the domain of \([ coul \]) to a second dimension. Let \( \mathbb{C} \) be the set of conformations over which an energy surface is defined. For each conformation \( x \in \mathbb{C} \), (1) charge density is taken from the wave function \( \psi_x \) without approximation; (2) for each pair of Gaussian primitives, the corresponding contribution to the electron density is assigned to a nucleus and is considered to be a part of that atomic density; (3) for each \( y \in \mathbb{C} \), atomic charge densities are translated and rotated for consistency with nuclear coordinates; and (4) \([ coul \])\(_x\)(\(y\)) is defined to be the electrostatic energy of the resulting molecular charge density. \([ coul \]) is an extension of \([ coul \]) in the sense that

\[
[ coul ](y) = [ coul ]_y(y)
\]  

An estimate of \([ elec \]) is obtained using

\[
[ elec ](y) = \left( \frac{1}{\# \mathbb{C}} \right) \sum_{x \in \mathbb{C}} [ coul ]\_x(y)
\]  

where \( \# \mathbb{C} \) is the number of conformations in \( \mathbb{C} \).

To obtain estimates of \([ p \]) and \([ r \]) (the \([ coul \]) subcomponents of \([ pol \]) and \([ rep \] + \([ tor \]), respectively), we introduce a new function \([ p \]) such that

\[
[ coul ]\_x(y) = [ elec ](y) + [ p ]\_x(y) + [ r ]\_x
\]  

\([ p \]) is an extension of \([ p \]) in the sense that

\[
[ p ](y) = [ p ]_y(y)
\]  

An estimate of \([ r \]) is obtained as follows. For each conformation \( x \in \mathbb{C} \), \([ r ]\) is chosen to minimize

\[
\sum_{y \in \mathbb{C}} \left( [ coul ]\_y(y) - ([ elec ](y) + [ r ]\_x) \right)^2
\]  

An estimate of \([ p \]) follows from eqs. (6) and (7). \([ p ](y) \) is a measure of the decrease in the electrostatic energy of the molecular charge density that results from deformation of atomic charge densities from their idealized unpolarized states in response to the external electric fields encountered in conformation \( y \).

Later in the Methods section, we show that \([ p \]) can be large relative to \( ab \) \( initio \) energy differences between conformations. This causes difficulty in obtaining an optimal estimate of \([ elec \]). We also show that \([ p \]), although large, is largely cancelled by \([ \tau ] + [ \epsilon ]_1 \), so that

\[
[ p ] + [ \tau ] + [ \epsilon ]_1 \ll [ p ]
\]  

In other words, the \([ coul \]) subcomponent of polarization is largely balanced by the \([ kin \] + \([ exch \]) subcomponent. This allows us to obtain reasonable representations of \([ pm \]) using molecular mechanics models that fail to account for polarization or other flow of electron density with torsion angle rotation. Also, because the \([ elec \]) energy surface is just one component of a more general picture, errors in the representation of \([ pol ] + [ rep ] + [ tor ] \) complicate attempts to determine the relative accuracies of a collection of electrostatic models.

**AB INITIO ENERGY SURFACE**

The \((\phi, \psi)\) energy surface of blocked alanine (N-acetyl-N'-methyl alanineamide) was calculated at the HF/6-31G* level over a 24 \times 24 grid using the \( ab \) \( initio \) molecular orbital program SPARTAN.\(^\text{22}\) All bond lengths, bond angles, and torsion angles other than \( \phi \) and \( \psi \) were maintained rigid. Figure 1 introduces notation for the atoms of blocked alanine. The geometry of blocked alanine that was used, both here and in the molecular mechanics calculations, is specified in Table I. The bond lengths and bond angles of the peptide group are those suggested by Benedetti.\(^\text{23}\) Other elements of the geometry were taken from Momany et al.\(^\text{1}\)

In the region of the \( C_{\text{axial}} \) conformation, and possibly in other regions as well, molecular mechanics model energy surfaces can change sharply within the 15° intervals between grid points. Therefore, deviations between model and \( ab \) \( initio \) energy surfaces might conceivably become small at grid points while remaining large at points between. To guard against this possibility, the \( ab \) \( initio \) energy surface was interpolated to a larger \( 72 \times 72 \) grid, and our measure of distance between
two surfaces was defined to be sensitive to deviations between surfaces at all grid points of this larger grid. Roughly speaking, interpolation was accomplished by minimizing the curvature of the curves created by intersection of the surface with planes of constant $\phi$ or $\psi$.

The following is a more detailed description of the interpolation procedure. Let $H$ be the set of grid points of the larger $72 \times 72$ grid and $f_{i,j}$ the energy at grid point $(i,j)$. At the grid points of the smaller $24 \times 24$ grid, $f_{i,j}$ is held fixed. The remaining $f_{i,j}$ are adjusted to minimize

$$
\sum_{(i,j) \in H} \left( \frac{(f_{(i+1,j)} - 2f_{i,j} + f_{(i-1,j)})(f_{(i+1,j)} - f_{i,j})}{[1 + (f_{(i+1,j)} - f_{i,j})^2]^{1/2}} + \frac{(f_{(i,j+1)} - 2f_{i,j} + f_{(i,j-1)})(f_{(i,j+1)} - f_{i,j})}{[1 + (f_{(i,j+1)} - f_{i,j})^2]^{1/2}} + \frac{(f_{(i,j+1)} - 2f_{i,j} + f_{(i,j-1)})(f_{(i,j+1)} - f_{i,j})}{[1 + (f_{(i,j+1)} - f_{i,j})^2]^{1/2}} \right)
$$

(10)

Equation (10) was obtained as follows. Within a localized region of the $(\phi, \psi)$ plane that contains grid point $(i,j)$, the curve that is created by intersection of the ab initio energy surface with a plane of constant $\psi$ can be approximated by the quadratic polynomial

$$
F(\phi) = \frac{a}{2}(\phi - \phi_i)^2 + b(\phi - \phi_i) + c
$$

(11)

![Figure 1. Notation used to identify atoms of blocked alanine.](image)

**Table II.**

Geometry of Blocked Alanine in Z-matrix Format.

<table>
<thead>
<tr>
<th>Bond lengths (Å)</th>
<th>Bond angles (degrees)</th>
<th>Torsion angles (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C_2-C_1$</td>
<td>1.510</td>
<td></td>
</tr>
<tr>
<td>$N_2-C_2$</td>
<td>1.335</td>
<td></td>
</tr>
<tr>
<td>$C_4-N_3$</td>
<td>1.449</td>
<td>116.70</td>
</tr>
<tr>
<td>$C_4-C_4$</td>
<td>1.522</td>
<td>121.90</td>
</tr>
<tr>
<td>$N_4-C_4$</td>
<td>1.335</td>
<td>109.30</td>
</tr>
<tr>
<td>$C_6-N_5$</td>
<td>1.449</td>
<td>116.70</td>
</tr>
<tr>
<td>$C_6-C_6$</td>
<td>1.530</td>
<td>121.90</td>
</tr>
<tr>
<td>$H_9-C_9$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
<tr>
<td>$H_{10}-C_1$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
<tr>
<td>$H_{11}-C_1$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
<tr>
<td>$O_{12}-C_{12}$</td>
<td>1.229</td>
<td>109.47</td>
</tr>
<tr>
<td>$H_{13}-N_3$</td>
<td>1.090</td>
<td>119.70</td>
</tr>
<tr>
<td>$H_{14}-C_4$</td>
<td>1.090</td>
<td>119.70</td>
</tr>
<tr>
<td>$O_{15}-C_{15}$</td>
<td>1.229</td>
<td>107.73</td>
</tr>
<tr>
<td>$H_{16}-N_5$</td>
<td>1.090</td>
<td>107.73</td>
</tr>
<tr>
<td>$H_{17}-C_7$</td>
<td>1.090</td>
<td>119.70</td>
</tr>
<tr>
<td>$H_{18}-N_7$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
<tr>
<td>$H_{19}-C_9$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
<tr>
<td>$H_{20}-C_8$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
<tr>
<td>$H_{21}-C_8$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
<tr>
<td>$H_{22}-C_8$</td>
<td>1.090</td>
<td>109.47</td>
</tr>
</tbody>
</table>
where

\[ a = (f_{i+1,j} - 2f_{i,j} + f_{i-1,j}) \]
\[ b = \frac{(f_{i+1,j} - f_{i-1,j})}{2} \]
\[ c = f_{i,j} \]

and where the unit of circular arc that is used to measure \( \phi \) is taken to be the 5° separation between grid points. Over the segment of this curve that is defined by the domain \([\phi_0, \phi_1, \phi_2, \ldots, \phi_n]\), the line integral of the square of the curvature is

\[ \int (\phi - \frac{a}{\sqrt{2}}) \left( 1 + \left( \frac{a}{2} + b \right)^2 \right)^{1/2} d\phi \]

The first half of eq. (10) is obtained from eq. (15) by expressing \( a \) and \( b \) in terms of energies at grid points. In most regions, interpolation is justified by the smoothness of the \( ab \) initio energy surface. However, even when the \( ab \) initio energy surface is chopped at 16 kcal/mol relative to the minimum energy, energy changes of greater than 5 kcal/mol do occur between adjacent points of the 24 \( \times \) 24 grid in the steepest regions. In these regions, the interpolated surface could conceivably deviate from the actual \( ab \) initio energy surface by \(-1\) kcal/mol. Because the area of these regions is small in comparison to the area of the grid, errors introduced by interpolation are not expected to alter results in any significant way.

Figure 2 is a contour plot of our \( ab \) initio energy surface for \((\phi, \phi)\) of blocked alanine. This surface will be used as a target with respect to which relative accuracies will be determined for a collection of molecular mechanics model energy surfaces.

The accuracy of our HF/6-31G* energy surface can be inferred from two recent \( ab \) initio studies of peptides. From calculations on small polar molecules, it is known that \( ab \) initio wave functions obtained at the HF/6-31G* level tend to overestimate molecular dipoles by 10 to 15%. Historically, this error has often been compensated for through the scaling of partial charges obtained from HF/6-31G* wave functions. A recent article by Price et al. looks closely at the effect of basis set and electron correlation on the predicted electrostatic potential of peptides. Price et al. conclude that, at the HF/6-31G* level, almost all of the remaining error originates from the lack of electron correlation, and about three quarters of this error can be recovered by a scaling of the charge density. This result suggests that a simple scaling of our \( ab \) initio energy surface might also recover much of the remaining error. The effectiveness of scaling energies would depend on the strength of the correlation between errors in energy differences between conformations and errors in electron density. This correlation is expected to be good for the electrostatic component but is not known for other components, such as repulsion or intrinsic torsional. The recent work of T. Head-Gordon et al. suggests that the largest changes to energy differences between conformations that result from addition of electron correlation are \(-1.5\) kcal/mol. If the majority of this error could be recovered by a scaling of energies, then for most energy differences a conservative estimate of the remaining error would be less than .50 kcal/mol.

All of our electrostatic models are obtained directly from the HF/6-31G* wave functions, whose energies are used as the target with respect to which the relative accuracies of the electrostatic
models are determined. Therefore, any conclusions concerning the relative accuracies of our electrostatic models would not be altered by the substitution of wave functions and energy surfaces obtained using larger basis sets and electron correlation.

ELECTROSTATIC MODELS

Table III lists the electrostatic models that will be examined in this work. These models differ with respect to (1) restrictions placed on the form of atomic charge densities and (2) approximations used in the representation of electrostatic interaction energies between pairs of atomic charge densities. They range in complexity from crude to extremely detailed. Consistent with our definition of the [elec] component of the ab initio energy, the electrostatic models of Table III make no attempt to account for polarization or other flow of electron density with conformation.

In Table III, we distinguish three types of electrostatic model based on the approximations that are introduced. In partial charge models, atomic charge densities are required to be spherically symmetric, and the representation of electrostatic interaction energies between pairs of atomic charge densities neglects overlap. In multipole models, the requirement that atomic charge densities be spherically symmetric is relaxed, but the multipole-expansion representation of electrostatic interaction energies continues to neglect overlap of atomic charge densities. In nonspherical CPK models, no restriction is placed on the symmetry of atom charge densities, and no approximation is made in the calculation of electrostatic interaction energies between these densities. Most of the current generation of protein potential functions use a partial charge electrostatic model.1-4,6,7

Electrostatic models that allow nonspherical atomic charge densities predict that 1-2 interactions can depend on one torsion angle degree of freedom and that 1-3 interactions can depend on two torsion angle degrees of freedom. In contrast, a partial charge model predicts that electrostatic interaction energies of types 1-2 and 1-3 are independent of torsion angle rotation. For example, in blocked alanine, if dipoles, in addition to monopoles, are included in the representation of atomic charge densities, then the electrostatic interaction energy between N\(_{α}\) and C\(_{α}\) should change as a function of both \(ϕ\) and \(ψ\). For models \([\text{md}]\), \([\text{mdq}]\), \([\text{mdqo}]\), and \([\text{cpk}]\), electrostatic interaction energies are calculated for all pairs of atomic charge densities.

It is hard to see how short-range electrostatic interaction energies between nonspherical atomic charge densities could be accurately represented by any partial charge model. However, errors introduced by a poor representation of the electrostatic component might be compensated for by offsetting errors in other components. For example, errors in the representation of electrostatic energy that are dependent on only one torsion angle degree of freedom could easily be picked up by the one-dimensional Fourier terms that are commonly used to represent intrinsic torsional energy. Of greater concern are errors in the representation of

<table>
<thead>
<tr>
<th>Model</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>([\text{mul}])</td>
<td>Partial charge</td>
<td>Mulliken partial charges at nuclei Potential derived partial charges at nuclei</td>
</tr>
<tr>
<td>([\text{pd}])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>([\text{md}])</td>
<td>Multipole(^a)</td>
<td>A monopole and dipole are included at each nucleus</td>
</tr>
<tr>
<td>([\text{mdq}])</td>
<td></td>
<td>A monopole, dipole, and quadrupole are included at each nucleus</td>
</tr>
<tr>
<td>([\text{mdqo}])</td>
<td></td>
<td>A monopole, dipole, quadrupole, and octopole are included at each nucleus</td>
</tr>
<tr>
<td>([\text{cpk}])</td>
<td>Nonspherical CPK(^a)</td>
<td>Generalized atomic charge densities represented without approximation</td>
</tr>
</tbody>
</table>

\(^a\)Includes electrostatic interaction energies of types 1—2 and 1—3.
electrostatic energy that depend simultaneously on
two torsion angle degrees of freedom. It is hard to
see how these errors could be picked up using
standard functional forms.

The electrostatic energy surfaces of the partial
charge models [mul] and [pd] were obtained as
follows. For each of the $24 \times 24$ wave functions
whose energies define our target $ab$ initio energy
surface for $(\phi, \psi)$ of blocked alanine, two sets of
partial charges were calculated: Mulliken\textsuperscript{27} and
potential derived.\textsuperscript{28} In an attempt to remove polar-
electrostatic energy that depend simultaneously on
two torsion angle degrees of freedom. It is hard to
charge models

surface for
energies. The electrostatic energy surfaces of the
models [mul] and [pd] were calculated using these
averaged sets of partial charges over a $72 \times 72$
grid. For the 59 low-energy conformations that
were used in averaging, changes with conforma-
tion are larger by about a factor of 3 for potential
derived partial charges than for Mulliken partial
charges.

The electrostatic energy surfaces of the multi-
pole models [md], [mdq], and [mdqo], where m,
d, q, and o indicate which moments (monopole,
dipole, quadrupole, and octopole) are included at
each nucleus, were obtained as follows. Dis-
distributed atomic multipoles\textsuperscript{15} were calculated from
the wave functions whose energies define our tar-
get $ab$ initio energy surface for $(\phi, \psi)$ of blocked
alanine. For each atom and for each conformation,
the atomic multipoles were expressed with respect
to a local coordinate system defined by the posi-
tions of the nucleus and its connected nuclei.\textsuperscript{26}
To remove polarization contributions, atomic multi-
poles were averaged over the same set of 59 con-
formations that was used to obtain the averaged
partial charges of [mul] and [pd]. The electrostatic
energy surfaces of the models [md], [mdq], and
[mdqo] were calculated using this averaged set of
atomic multipoles over a $72 \times 72$ grid. For each
atom and for each conformation of the grid, the
averaged atomic multipoles were translated and
rotated along with the fragment defined by the
nucleus and its connected nuclei. For each pair of
atomic charge densities, the exact electrostatic in-
teraction energy was approximated using a trun-
cated multipole expansion.\textsuperscript{17} Computer programs
for (1) calculating atomic multipoles from wave
functions, (2) expressing atomic multipoles with
respect to a local coordinate system, (3) translating
and rotating atomic multipoles, and (4) calculating
multipole–multipole interaction energies were de-
veloped by the authors.

The product of two Gaussians is itself a Gauss-
ian, the center of the product lying on the line
connecting the centers of the two factors. In what
follows, we refer to the center of a product of two
Gaussian primitives as a GG site and to the nuclei
at the centers of the two Gaussian primitive factors
as parent nuclei.

Stone's algorithm for obtaining atomic multi-
poles from wave functions\textsuperscript{18} consists of the follow-
ing steps: (1) Multipoles are calculated at GG sites
for elements of electron density corresponding to
products of Gaussian primitives, and (2) multi-
poles are moved from GG sites to the nearest
nucleus. If a GG site is equidistant from two or
more nuclei, then equal contributions are moved
to each. A multipole is moved by replacing it with
the infinite sequence of multipoles that would
have been obtained from the element of charge
density that produced it had the integration that
relates an element of density to its multipole mo-
ments used a different origin.\textsuperscript{17} Here, the primary
consideration in the assignment of elements of
electron density to nuclei is rapid convergence of
the resulting multipole-expansion representation
of electrostatic energies.

In this article, the goal of multipole expansion—accurate representation of the elec-
 trostatic energy of a molecule that contains torsion
angle degrees of freedom—differs slightly from
the goal of multipole expansion in most previous
work.\textsuperscript{15} Consequently, the primary considera-
tion in the assignment of elements of electron density
to nuclei must be the movement with conforma-
tion that follows from this assignment. For this
reason, our algorithm for the assignment of ele-
ments of electron density to nuclei, described in
Table IV, differs somewhat from Stone's.

The electrostatic energy surface of the model
[cpk] was obtained as follows. For each of the 59
low-energy conformations that were used to obtain
averaged partial charges and multipoles, a set of
atomic charge densities was obtained without ap-
proximation from the corresponding HF/6-31G* wave function. These sets of atomic charge densi-
ties were used to calculate 59 electrostatic energy
surfaces over a $24 \times 24$ grid. For each set of atomic
charge densities and for each conformation of the
$24 \times 24$ grid, the atomic charge density of each
atom was translated and rotated along with the
TABLE IV.
Assignment of GG Sites\(^a\) to Nuclei\(^b,c\) for Electrostatic Models [md], [mdq], and [mdqo].

<table>
<thead>
<tr>
<th>Range of parent nuclei</th>
<th>Assigned nucleus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 — 1</td>
<td>Identical parent nuclei</td>
</tr>
<tr>
<td>1 — 2</td>
<td>Nearest nucleus</td>
</tr>
<tr>
<td>1 — 3</td>
<td>If a GG site is equidistant from two (or more) nuclei,(^d) then equal contributions are moved to both (each)</td>
</tr>
<tr>
<td>1 — 4 and higher</td>
<td>Connecting nucleus</td>
</tr>
<tr>
<td></td>
<td>Nearest nucleus</td>
</tr>
<tr>
<td></td>
<td>If a GG site is equidistant from two (or more) nuclei,(^d) then equal contributions are moved to both (each)</td>
</tr>
</tbody>
</table>

\(^a\)The center of a product of two Gaussian primitives.
\(^b\)Multipole is calculated at GG sites for elements of electron density corresponding to products of Gaussian primitives and then moved\(^d\) from GG sites to the assigned nucleus.
\(^c\)The assignment of elements of electron density to nuclei determines the motion of these elements relative to the nuclei in the resulting electro-static models.

Fragment defined by its nucleus and the set of connected nuclei. For each pair of atomic charge densities, the electrostatic interaction energy was calculated without approximation using exact integration. Finally, the electrostatic energy surface of the model [cpk] was obtained by averaging these 59 energy surfaces and interpolating the averaged surface, defined over a 24 x 24 grid, to a 72 x 72 grid. Interpolation was accomplished using the procedure that was described earlier in the Methods section. Because of the smoothness of the electrostatic energy surface throughout the low-energy regions, interpolation is not expected to introduce significant errors.

A computer program, cpk, for (1) obtaining atomic charge densities from wave functions without approximation, (2) translating and rotating these atomic charge densities for consistency with the nuclear coordinates of other conformations, and (3) calculating electrostatic interaction energies between pairs of atomic charge densities using analytical integration was developed by the authors. The calculation of two-electron integrals is based on the two-center expansion of \(1/r\) in spherical harmonics, which was derived by R. Sack.\(^{30-32}\)

Atomic charge densities were obtained from the molecular electron density by assigning each element of electron density to a nucleus, as described in Table V. For cases in which the two parent nuclei of a GG site can change relative position or orientation as a function of torsion angles \(\phi\) or \(\psi\), the closest parent nucleus was decided by comparing the normalized distances \(r_{1a}/R_a\) and \(r_{1b}/R_b\), where \(r_{1a}\) is the distance between the GG site and atom \(a\) and \(R_a\) is the van der Waals radius of atom \(a\). Here, the van der Waals radii of C, N, O, and H\(_2\) were taken to be 1.90 Å, 1.75 Å, 1.60 Å, and 1.45 Å, respectively.

For each atomic charge density, multipoles were calculated up to order 7. Then, in the calculation of electrostatic energy surfaces, whenever the distance between two nuclei became greater than 7.2 Å, the exact electrostatic interaction energy between the corresponding pair of atomic charge densities was approximated using a multipole expansion truncated at order 7. This approximation introduces no significant error and speeds the calculations by about a factor of 2.

Alternatively, an electrostatic energy surface of the type nonspherical CPK could have been obtained by first averaging atomic charge densities and then using the set of averaged atomic charge densities to calculate an electrostatic energy surface. This would have been more difficult computationally because of the large storage capacity required by a properly averaged set of atomic charge densities. Presumably, the electrostatic energy surface of type nonspherical CPK obtained using this different method of averaging would be similar in its major features to the electrostatic energy surface of the model [cpk].

Because of the extreme detail with which atomic charge densities are represented, a nonspherical CPK model of blocked alanine becomes protein-like in its number of interaction sites. For blocked alanine, the number of interaction sites (the number of GG sites) is 12,561, of which about 3000 contribute significantly to electrostatic interaction energies. Although the time needed to compute electrostatic energies using a nonspherical CPK model remains an order of magnitude less than the time needed to compute \(ab\) \textit{initio} energies, this model is
TABLE V.
Assignment of GG Sites\(^a\) to Nuclei for Electrostatic Model (cpk)\(^b\)

<table>
<thead>
<tr>
<th>Range of parent nuclei</th>
<th>Can the two parent nuclei of a GG site change relative position or orientation as a function of $\phi$ or $\psi$?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 — 1</td>
<td>Identical parent nuclei</td>
</tr>
<tr>
<td>1 — 2</td>
<td>Closest parent nucleus</td>
</tr>
<tr>
<td>1 — 3</td>
<td>Connecting nucleus</td>
</tr>
<tr>
<td>1 — 4 and higher</td>
<td>Closest nucleus within the rigid fragment containing the two parent nuclei</td>
</tr>
</tbody>
</table>

\(^a\)The center of a product of two Gaussian primitives.
\(^b\)A set of atomic charge densities is obtained from the molecular electron density by assigning each element of electron density to a nucleus.

much too slow for use in typical molecular mechanics applications. In this article our primary motivation for introducing a nonspherical CPK model is to enable us to look at the effect—on distance between molecular mechanics model and \textit{ab initio} energy surfaces—of substituting a highly accurate representation for electrostatic interaction energies of types 1—2 and 1—3. In the future, a nonspherical CPK model might be used to parameterize a correction for errors that are introduced into estimates of electrostatic interaction energies of types 1—2 and 1—3 by the use of a more tractable multipole model.

MODELS OF REPULSION + DISPERSION + INTRINSIC TORSIONAL ENERGY

Table VI lists the models of repulsion + dispersion + intrinsic torsional energy that will be used in this work. The sum of molecular mechanics components is our definition of the molecular mechanics component $[\text{rep}] + [\text{tor}]$. The $7 \{ [\text{rep}] + [\text{tor}] \}$

```
TABLE VI.
{[rep]} + {[tor]} Models Used to Determine Relative Accuracies of Electrostatic Models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amber</td>
<td>Full atom amber(^2,36)</td>
</tr>
<tr>
<td>eccep</td>
<td>Modified eccep(^2,33)</td>
</tr>
<tr>
<td>opls4</td>
<td>The large 12–10 contributions to H bonds were removed opsl(^4)</td>
</tr>
<tr>
<td>charmm</td>
<td>charmm version 22(^2)</td>
</tr>
<tr>
<td>mm3</td>
<td>mm3(^2,24,25)</td>
</tr>
<tr>
<td>merck</td>
<td>merck molecular force field(^6)</td>
</tr>
<tr>
<td>decomp</td>
<td>Estimate of repulsion + intrinsic torsional energy obtained by recognizing patterns of change in the quantum mechanics components [kin], [coull] and [exchl] that are characteristic of repulsion and intrinsic torsional energy</td>
</tr>
</tbody>
</table>
```
[tor] models of Table VI are combined with the six electrostatic models of Table III to give 42 molecular mechanics models. We do not attempt to determine the relative accuracies of these seven [rep] + [tor] potential functions, which in some cases have been parameterized in combination with their electrostatic components and, therefore, have limited meaning when combined with other electrostatic models. By choosing a range of [rep] + [tor] models, our purpose is only to provide a representative sample of functional forms and parameter sets with which we can test our electrostatic models.

The [rep] + [tor] energy surfaces of the models [amber], [cepp], [opls], [charmm], [mm3], and [merck] were calculated over a 72 x 72 grid in the torsion angles $\psi$ and $\phi$. The model [cepp] is essentially cepp repulsion + dispersion. The large 12—10 contributions to H bonds, which would have been inconsistent when combined with electrostatic models that use a dielectric constant of 1, were removed. The [rep] + [tor] energy surfaces of the models [mm3], [merck], [amber], and [opls] are plotted in Figure 3. From Figure 3, we see that these energy surfaces differ widely among themselves. The model [decomp] was determined directly from the quantum mechanics energy components [kin], [coul], and [exch] and, therefore, includes both repulsion and intrinsic torsional energy, but not dispersion.

To find a pattern of change in the components [kin], [coul], and [exch] that correlates with an increase in [rep] + [tor], we looked at three changes of the conformations of three simplified systems, each expected to cause a pure change

![Contour plots of energy surfaces](image-url)
in one of three molecular mechanics components—repulsion, polarization, or intrinsic torsional energy. Then, for each of these three changes, we determined the corresponding pattern of change in the components \([\text{kin}], [\text{coul}], \text{ and } [\text{exch}]\).

As a model of a pure change to the intrinsic torsional component, we chose ethane and torsion angle rotation from staggered to eclipsed. Here, little change is expected in electrostatic, polarization, and repulsion components. In Table VII, the geometry is specified for a sequence of 19 conformations of ethane as a function of the torsion angle \(\chi\). The ab initio energy curve was calculated over these 19 conformations at the HF/6-31G* level. Components \([\text{kin}], [\text{coul}], \) and \([\text{exch}]\) were obtained using programs developed by the authors. The corresponding pattern of change in the quantum mechanics components \([\text{kin}], [\text{coul}], \text{ and } [\text{exch}]\), along with the ab initio energy curve and the electrostatic energy curve of the model \([\text{cpk}]\), are plotted in Figure 4.

As \(\chi\) changes from 60° to 0°, both \([\text{coul}]\) and \([\text{exch}]\) decrease. These decreases are offset by a sharp increase in \([\text{kin}]\). A possible physical origin for this pattern of change is suggested by the observed flow of electron density out of the center of the C—C bond. A buildup of electron density, relative to the density that results from superposition of spherically symmetric atomic charge densities, is predicted by quantum mechanics to occur at the center of each covalent bond. The extent of this buildup is the result of a compromise between the competing preferences of kinetic and potential energies. The potential energy \([\text{coul}] + [\text{exch}]\) is decreased by increases to the electron density in the neighborhood of nuclei and, therefore, tends to prefer a superposition of spherically symmetric atomic charge densities. In contrast, the kinetic energy is decreased by increases to the smoothness of the wave function and, therefore, tends to prefer buildup of electron density at bond centers. Based on these considerations, the observed pattern of change in \([\text{kin}], [\text{coul}], \text{ and } [\text{exch}]\) is possibly the result of flow of electron density out of the center of the C—C bond. Studies relevant to possible causes of this flow have been reviewed by Potter.\(^{16}\)

As expected, the electrostatic energy curve of the model \([\text{cpk}]\) is nearly flat. The predicted electrostatic barrier to rotation is 44 kcal/mol. This electrostatic energy curve is an average over 12 curves, obtained using the 12 sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with \(\chi = 0°\) to \(\chi = 110°\).
densities that were obtained without approximation from the 12 (distinct) wave functions whose energies define the \textit{ab initio} energy curve. Using the sets of atomic charge densities obtained from staggered and eclipsed wave functions, the corresponding barriers to rotation are .24 and .64 kcal/mol, respectively. This result suggests that, as expected, the polarization component is small. Although the contribution to energy differences between conformations from short-range electrostatic interactions is small in ethane, it is expected to be larger in blocked alanine, where low-order atomic multipole moments are larger.

A barrier to rotation in ethane of 3.62 kcal/mol is predicted by our \textit{ab initio} energy curve. This energy curve drops to -3.0 kcal/mol with the inclusion of geometry optimization and electron correlation. The barrier to rotation in ethane of 3.62 kcal/mol is predicted by our \textit{ab initio} energy curve. This value drops to -3.0 kcal/mol with the inclusion of geometry optimization and electron correlation. The component curves [kin], [coul], and [exch] change considerably when bond lengths and bond angles are relaxed.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

![Pattern of change in the quantum mechanics components (kin), (coul), and (exch)](figure)

Corresponding to the change to the conformation of the ethane-ethane dimer that is specified in Table VIII. Also included are the \textit{ab initio} energy curve and the electrostatic energy curve of the model [cpk]. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

![Pattern of change in the quantum mechanics components (kin), (coul), and (exch)](figure)

Corresponding to the change to the conformation of the ethane-ethane dimer that is specified in Table VIII. Also included are the \textit{ab initio} energy curve and the electrostatic energy curve of the model [cpk]. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.

As a model of a pure change to the repulsion component, we chose the ethane-ethane dimer and a decrease in the distance between methyl groups from 6.6 to 3.0 Å. Here, little change is expected in electrostatic, polarization, and intrinsic torsional components. In Table VIII, the geometry is specified for a sequence of 19 conformations of the ethane-ethane dimer as a function of the distance R between carbon nuclei. The \textit{ab initio} energy curve was calculated at the HF/6-31G* level. The electrostatic energy curve is the average of eight curves, obtained using the eight sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with R = 5.2 Å to R = 6.6 Å.
As \( K \) decreases from 6.6 to 3.0 Å, both [\text{coul}] and [\text{exch}] decrease. These decreases are offset by a sharp increase in [\text{kin}]. Qualitatively, this pattern of change is similar to the pattern observed in our model of pure change to the intrinsic torsional component. The primary physical origin of this pattern of change is probably contraction of atomic charge densities (to reduce volumes of overlap) forced by the requirement that molecular orbitals be orthogonal. Contraction would be expected to alter the balance between kinetic and potential energies, decreasing potential at the expense of kinetic. A secondary physical origin is probably penetration of atomic charge densities, which would be expected to increase the corresponding electrostatic interaction energies.

As a model of a pure change to the polarization component, we chose the water-ethane dimer and a change in the relative orientation of the two molecules separated by a distance of 6.00 Å. Here, little change is expected in electrostatic repulsion, and intrinsic torsional components. In Table IX, the geometry is specified for a sequence of 19 conformations of the water-ethane dimer as a function of the angle \( \theta \) between the plane of the water molecule and the axis of the C-C bond. The ab initio energy curve was calculated over these 19 conformations at the HF/6-31G* level. The corresponding pattern of change in the quantum mechanics components [\text{kin}], [\text{coul}], and [\text{exch}], along with the ab initio energy curve and the electrostatic energy curve of the model [\text{cpk}], are plotted in Figure 6.

As \( \theta \) changes from 0° to 180°, the decrease in [\text{coul}] is largely offset by an increase in [\text{kin}]. The [\text{exch}] curve is essentially flat. This compensation between the [\text{coul}] and [\text{kin}] components is consistent with the large [\text{p}] component that is observed for blocked alanine (the large electrostatic favoring of the conformation from which a set of atomic charge densities was obtained). Also, differences in

<table>
<thead>
<tr>
<th>Bond lengths (Å)</th>
<th>Bond angles (Å)</th>
<th>Torsion angles (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{C}_2-\text{C}_1 )</td>
<td>1.530</td>
<td>( \text{H}_1-\text{C}_1-\text{C}_2 )</td>
</tr>
<tr>
<td>( \text{H}_2-\text{C}_1 )</td>
<td>1.090</td>
<td>( \text{H}_2-\text{C}_1-\text{C}_2 )</td>
</tr>
<tr>
<td>( \text{H}_3-\text{C}_1 )</td>
<td>1.090</td>
<td>( \text{H}_3-\text{C}_1-\text{C}_2 )</td>
</tr>
<tr>
<td>( \text{C}_2-\text{C}_1 )</td>
<td>1.530</td>
<td>( \text{H}_1-\text{C}_1-\text{C}_2 )</td>
</tr>
<tr>
<td>( \text{H}_1-\text{O}_1 )</td>
<td>1.00</td>
<td>( \text{H}_1-\text{O}<em>1-\text{X}</em>{10} )</td>
</tr>
<tr>
<td>( \text{H}_2-\text{O}_2 )</td>
<td>0.889/41</td>
<td>( \text{H}_2-\text{O}<em>2-\text{X}</em>{10} )</td>
</tr>
</tbody>
</table>

*\( \theta \) is 0°, 10°, 20°, . . . , 180°.

TABLE IX.
A Sequence of Conformations for the Water-Ethane Dimer (in Z-Matrix Format) over Which Changes in the Molecular Mechanics Components of the Energy Are Expected to be Concentrated in [pol].
[pol] = [p] + [τ] + [ε], between conformations are much smaller than differences in [p]. This supports our earlier assumption that the [coul] component of polarization is probably largely balanced by the [kin] + [exch] component. Because changes in [rep] + [tor] and [pol] are negligible and small, respectively, [elec] is a good approximation of [qm].

Table X summarizes the patterns of change in [kin], [coul], and [exch] that result from changes in the conformations of our three simplified systems and correspond to pure increases in one of three molecular mechanics components. These patterns allow us to define a crude quantum mechanics-based [rep] + [tor] model that is independent of our previously defined [elec] models.

To construct a function that maps changes in the quantum mechanics components [kin], [coul], and [exch] into changes in the molecular mechanics component [rep] + [tor], we first looked at the pattern of change in [kin], [coul], and [exch] that results from a change in [rep] + [tor] and attempted to isolate from this pattern some characteristic element that could not be hidden by the superposition of patterns that result from changes in [elec] or [pol]. Because changes in the [kin] and [coul] components of [pol] tend to become large and offset one another, and because changes in the [exch] component of [pol] remain small, we chose to focus on [exch]. For ethane, the ratio of change in intrinsic torsional energy to change in [exch] is relatively constant at approximately −1.68. More formally, this relationship can be expressed as

\[ \text{[rep]} + \text{[tor]} \approx (-1.68)(\text{[exch]} - \text{[exch]}_\text{max}) \]

(16)

where [exch]_max is the maximum of the [exch] curve at \( \chi = 60^\circ \). For the ethane–ethane dimer, the ratio of change in repulsion to change in [exch] is approximately −0.7 when repulsion is small but increases slightly as repulsion increases. This relationship can be expressed more formally as

\[ \text{[rep]} + \text{[tor]} = (-0.700)(\text{[exch]} - \text{[exch]}_\text{max}) + (0.16)(\text{[exch]} - \text{[exch]}_\text{max})^2 \]

(17)

where [exch]_max is the maximum of the [exch] curve and where the unit of energy is kcal/mol.

For blocked alanine, the [rep] + [tor] energy surface of the model [decomp], defined over a 72 × 72 grid in torsion angles \( \phi \) and \( \psi \), was obtained as follows. The quantum mechanics components [coul], [exch], and [kin] of our ab initio energy surface were calculated from the wave functions over a 24 × 24 grid. The [rep] + [tor] energy surface of the model [decomp] was defined over a 24 × 24 grid to be the quadratic function of the [exch] surface that is specified by the right-hand side of eq. (17), where [exch]_max is taken to be the maximum of the [exch] surface. Finally, this surface, defined over a 24 × 24 grid, was interpolated to a larger 72 × 72 grid using the procedure that was used to interpolate the ab initio energy surface. The [rep] + [tor] energy surface of the model [decomp] is plotted in Figure 7.

The fact that eq. (16) differs from eq. (17) suggests that this [decomp] estimate of [rep] + [tor] is far from optimal. With further study, a more accurate definition of [decomp] in terms of [coul], [exch], and [kin] should be possible.

**CONVERGENCE OF 1–2 AND 1–3 MULTIPOLE–MULTIPole INTERACTION ENERGIES**

In this section, we look at errors introduced into calculations of electrostatic interaction energies between pairs of atomic charge densities by the use of a truncated multipole expansion. This is one of two major problems that complicate attempts to use Stone's distributed multipole analysis (dma) to represent the electrostatic energy of a flexible molecule. The second problem, the unwanted contribution to electrostatic energy differences between conformations that results from polarization-induced deformations of atomic charge densities from their idealized unpolarized states, will be discussed in the following section. Together, these two sections summarize the preliminary results, conclusions, and logic that led us to the following decisions concerning our three multipole electrostatic models: (1) the decision not to include interaction sites at bond centers and other nonnuclear positions, (2) the decision to include interaction
energies of types 1—2 and 1—3, and (3) the decision to average atomic charge densities over a collection of conformations.

Let \( r_1 \) and \( r_2 \) be the maximum extents of two atomic charge densities \( p_1 \) and \( p_2 \) from the positions of their nuclei \( a_1 \) and \( a_2 \). Derivation of the multipole-expansion representation of the electrostatic interaction energy requires that \( r_1 + r_2 < |a_1 - a_2| \). If this condition is not met, then the multipole—multipole interaction energy will either fail to converge or lack physical meaning. Interactions of type 1—2 (distance \( \leq 1.5 \) \( \AA \)) occur inside of the surfaces of the bonded atomic charge densities (bygen 1.60 \( \AA \) nitrogen—carbon " 10.90 \( \AA \)). Interactions of type 1—3 (distance \( \leq 2.5 \) \( \AA \)) occur inside of the sum of the radii of the interacting atomic charge densities. Interactions of type 1—4 (distance \( \leq 2.9 \), 3.8 \( \AA \)) can occur slightly inside of the sum of the radii of the interacting atomic charge densities, but errors resulting from this overlap are probably negligible.

For both ethane and blocked alanine, we looked at the convergence of multipole-expansion representations of electrostatic interaction energies between pairs of atomic charge densities. For 1—2 interactions, multipole—multipole interaction energies fail to converge through order 7—order 7; they are too close. For 1—3 interactions, multipole—multipole interaction energies fail to converge through hexadecapole—hexadecapole. For 1—4 and higher interactions, reasonable convergence (less than 0.5 kcal/mol) can be obtained, but only if multipole moments through octopole are included at each nucleus.

For ethane, using atomic multipoles obtained from the staggered wave function, the electrostatic barrier to rotation through order 7—order 7 is 4.16 kcal/mol but has not yet converged. For comparison, a more exact estimate of the electrostatic barrier to rotation, obtained using a nonspherical CPK electrostatic model and atomic charge densities obtained from the staggered wave function, is 0.24 kcal/mol. This is an example of the error that can be introduced when a multipole-expansion representation of 1—2 electrostatic interaction energies is not truncated at low order.

If 1—3 multipole—multipole interaction energies could be caused to converge, it would remove the need for any two-dimensional corrections to the electrostatic energy of a multipole electrostatic model. In an attempt to obtain a better representation of 1—2 and 1—3 electrostatic interaction energies, we explored the effect of including interaction sites at bond centers and at the midpoints of 1—3 type atom pairs. Success was measured by the amount that higher order multipole moments were reduced and by convergence of 1—3 multipole—multipole interaction energies. We find that the magnitudes of higher order multipole moments can be reduced by increasing the number of interaction sites. However, 1—3 multipole—multipole interaction energies do not converge, regardless of the choice of interaction sites. This is because distances between interaction sites decrease as the number increases.

Lack of convergence and physical meaning of multipole—multipole interaction energies at bond distances caused us to look at the effect of neglecting 1—2 interactions in multipole electrostatic models. For the multipole electrostatic model \([md]\), distances between model and ab initio energy surfaces indicate that inclusion of all interactions is preferable to inclusion of only 1—3 and higher interactions. In contrast, for models \([mdq]\) and \([mdq0]\), the accuracies increase and remain approximately unchanged, respectively, when 1—2 interactions are excluded. This result suggests that 1—2 interactions between lower order multipole moments have greater physical meaning than 1—2 interactions between higher order moments.

EFFECTS OF POLARIZATION-INDUCED DEFORMATION OF ATOMIC CHARGE DENSITIES

In this section, we look at the effect of polarization on electrostatic energies using both multipole and nonspherical CPK models. The multipole
models allow easy resolution with respect to range and moment–moment contributions but are burdened by lack of physical meaning and convergence at short range. To simplify the following discussion, the conformation from which a set of atomic multipoles or charge densities was obtained will be referred to as the incumbent conformation.

Let $A$ be the set of atoms in a molecule. For each conformation $x \in C$ of a polar molecule, the corresponding set of atomic charge densities $\{p_a : a \in A\}$ is slightly deformed relative to the set of idealized unpolarized atomic charge densities $\{p_a : a \in A\}$. For each atom $a \in A$ the small deformation $p_a = (p_a - \langle p_a \rangle)$ is a response to external electric fields created by $\{p_b : b \in A, b \neq a\}$ in conformation $x$. The extended energy surface is obtained by translation and rotation of $\{p_a : a \in A\}$ with conformation according to the nonspherical CPK model. Here, $\langle p_a \rangle$ is the contribution to energy differences between conformations that results from the polarization $\langle p_a \rangle$ of atomic charge densities. Without exception, $\langle p_a \rangle$ favors the incumbent conformation $x$.

For $(\phi, \psi)$ of blocked alanine, the magnitude of this electrostatic favoring of the incumbent conformation is shown qualitatively in Figure 8 and quantitatively in Table XI. From Table XI, we see that $\langle \text{coul} \rangle(E)$ is lower than $\langle \text{coul} \rangle(A)$, that $\langle \text{coul} \rangle(A)$ is lower than $\langle \text{coul} \rangle(E)$, and that the energy difference between alpha helical and extended conformations changes by 20.6 kcal/mol depending on whether we lock in the density of the extended conformation or the density of the alpha conformation. A second measure of electrostatic favoring of the incumbent conformation is $\langle \text{coul} \rangle(x)$, the $\langle \text{coul} \rangle$ component of $\langle \text{pol} \rangle$. Table XII shows that, for the 59 conformations of a $12 \times 12$ grid whose ab initio energies are within 12 kcal/mol of the lowest, $\langle \text{coul} \rangle(x)$ is large. Clearly, for any single conformation $x$, $\langle \text{coul} \rangle$ is a poor estimate of $\langle \text{elec} \rangle$.

Because $\langle \text{coul} \rangle$ is so large, it is not obvious that a simple averaging of atomic charge densities or electrostatic energy surfaces should be an effective method for removing error caused by polarization from an estimate of $\langle \text{elec} \rangle$. For each conformation $x \in C$, where $C$ is the set of 59 conformations for which values of $\langle \text{coul} \rangle(x)$ are presented in Table XII, the ratio of the average absolute value of $\langle \text{coul} \rangle$ defined by

$$
\frac{1}{2C-1} \sum_{y \in C} |\langle \text{coul} \rangle(y)|
$$

(19)

to $-\langle \text{coul} \rangle(x)$ lies between .15 to .42, indicating a spike of stability at $x$. This result is consistent with a physical origin in which the many small electrostatic interaction energies that collectively form $\langle \text{coul} \rangle(y)$ lose coherence, tending to add less constructively, as $y$ moves away from $x$. The shape

\[\text{FIGURE 8. Electrostatic favoring of the incumbent conformation. Contour plots of model energy surfaces for (}\phi, \psi)\text{ of blocked alanine obtained by combining the [rep] + [tor] energy surface of the model [decomp] with [coul] energy surfaces calculated for incumbent conformations (a) (} -150, 150) and (b) (} -90, -30) using a nonspherical CPK model. The contour levels range from 1 to 16 kcal/mol in increments of 1 kcal/mol.}\]
TABLE XI. \[\text{[coul]}_{\phi, \psi}(\phi_2, \psi_2)\) Obtained Using a Nonspherical CPK Electrostatic Model for Nine Low-Energy Conformations of Blocked Alanine.\(^a\)

<table>
<thead>
<tr>
<th>(\phi_2, \psi_2)</th>
<th>A</th>
<th>G</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
<th>E</th>
<th>A*</th>
<th>C*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(−75, −45)</td>
<td>A</td>
<td>24.55</td>
<td>24.39</td>
<td>17.04</td>
<td>33.12</td>
<td>38.19</td>
<td>49.14</td>
<td>48.95</td>
<td>31.69</td>
</tr>
<tr>
<td>(−150, −55)</td>
<td>G</td>
<td>31.10</td>
<td>18.35</td>
<td>18.63</td>
<td>36.46</td>
<td>34.34</td>
<td>53.49</td>
<td>43.55</td>
<td>30.82</td>
</tr>
<tr>
<td>(−95, 10)</td>
<td>B</td>
<td>19.42</td>
<td>17.99</td>
<td>10.18</td>
<td>25.73</td>
<td>29.36</td>
<td>43.60</td>
<td>40.99</td>
<td>23.58</td>
</tr>
<tr>
<td>(−85, 75)</td>
<td>C</td>
<td>17.24</td>
<td>17.15</td>
<td>6.17</td>
<td>16.67</td>
<td>22.94</td>
<td>35.34</td>
<td>35.41</td>
<td>17.89</td>
</tr>
<tr>
<td>(−150, 80)</td>
<td>D</td>
<td>27.61</td>
<td>18.60</td>
<td>14.72</td>
<td>26.95</td>
<td>25.05</td>
<td>46.56</td>
<td>36.91</td>
<td>22.98</td>
</tr>
<tr>
<td>(−80, 150)</td>
<td>F</td>
<td>25.54</td>
<td>24.50</td>
<td>16.45</td>
<td>27.87</td>
<td>33.82</td>
<td>40.71</td>
<td>40.00</td>
<td>27.06</td>
</tr>
<tr>
<td>(−150, 150)</td>
<td>E</td>
<td>26.96</td>
<td>16.16</td>
<td>13.90</td>
<td>26.77</td>
<td>25.44</td>
<td>41.47</td>
<td>30.75</td>
<td>22.25</td>
</tr>
<tr>
<td>(60, 45)</td>
<td>A*</td>
<td>20.45</td>
<td>23.50</td>
<td>15.98</td>
<td>29.26</td>
<td>35.45</td>
<td>47.32</td>
<td>41.99</td>
<td>19.47</td>
</tr>
<tr>
<td>(60, −100)</td>
<td>C*</td>
<td>23.01</td>
<td>18.00</td>
<td>12.07</td>
<td>24.86</td>
<td>27.12</td>
<td>40.84</td>
<td>35.04</td>
<td>18.86</td>
</tr>
</tbody>
</table>

\(^a\)In units of kcal / mol.

that is observed for \([p]_z\), supports the expectation that error introduced by polarization into an estimate of \([\text{elec}]\) should be reduced by averaging.

Table XII shows that changes with conformations in \([p]_z(\phi)\) are certainly not negligible. However, as expressed in (9), we expect that changes in \([p]\) will, to a large extent, be balanced by opposite changes in \([\tau_1] + [\epsilon_1]\), causing great moderation in the molecular mechanics component \([\text{pol}]\). In molecular mechanics, the component \([\rho_1] + [\epsilon_1]\) corresponds to energy required to deform unpolarized atomic charge densities. Furthermore, we expect that changes in \([\text{pol}]\) with conformation will be small in comparison to the errors remaining in standard models of \([\text{elec}]\) and \([\text{rep}] + [\text{tor}]\). In other words, we expect that substituting increasingly accurate representations of electrostatic energy will reduce the distance between model and \(ab \text{ initio}\) energy surfaces, even within the context of a molecular mechanics model that neglects polarization.

For blocked alanine, a collection of "multipole" electrostatic models was used to resolve the \((\text{[coul]}_A(E) − \text{[coul]}_A(A)) − (\text{[coul]}_E(E) − \text{[coul]}_E(A))\) measure of electrostatic favoring of the incumbent conformation with respect to range and moment–moment contributions. The results, presented in Table XIII, indicate that \([p]_z\) favors the incumbent conformation at all levels and that this favoring tends to increase as the range is extended or the number of moment–moment contributions is increased. Consistent with this energetic picture,

TABLE XII. \([p]_{\phi_2, \psi_2}(\phi, \psi)\) Obtained for Blocked Alanine Using a Nonspherical CPK Electrostatic Model.\(^a,b,c\)

<table>
<thead>
<tr>
<th>(\psi)</th>
<th>−180</th>
<th>−150</th>
<th>−120</th>
<th>−90</th>
<th>−60</th>
<th>−30</th>
<th>0</th>
<th>30</th>
<th>60</th>
<th>90</th>
<th>120</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>−6.8</td>
<td>−4.5</td>
<td>−3.2</td>
<td>−4.66</td>
<td>−7.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>−6.4</td>
<td>−4.2</td>
<td>−2.7</td>
<td>−3.9</td>
<td>−7.5</td>
<td>−8.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>−6.0</td>
<td>−4.1</td>
<td>−3.2</td>
<td>−4.5</td>
<td>−6.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>−6.2</td>
<td>−4.7</td>
<td>−4.1</td>
<td>−6.1</td>
<td></td>
<td>−9.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>−6.5</td>
<td>−5.2</td>
<td>−3.9</td>
<td>−5.1</td>
<td></td>
<td>−7.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>−5.1</td>
<td>−4.1</td>
<td>−4.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−30</td>
<td>−6.9</td>
<td>−6.2</td>
<td>−5.5</td>
<td>−5.6</td>
<td>−7.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−60</td>
<td>−6.0</td>
<td>−5.4</td>
<td>−5.3</td>
<td>−6.0</td>
<td>−7.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−90</td>
<td>−5.8</td>
<td>−4.8</td>
<td>−4.7</td>
<td>−5.8</td>
<td></td>
<td>−3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−120</td>
<td>−6.4</td>
<td>−4.7</td>
<td>−4.5</td>
<td>−5.4</td>
<td></td>
<td>−6.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−150</td>
<td>−7.0</td>
<td>−4.9</td>
<td>−4.3</td>
<td>−5.6</td>
<td></td>
<td>−8.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−180</td>
<td>−7.0</td>
<td>−4.9</td>
<td>−3.8</td>
<td>−5.3</td>
<td>−8.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)[p]_{\phi_2, \psi_2}(\phi, \psi) is a measure of electrostatic favoring of the conformation from which a set of atomic charge densities was obtained.

\(^b\)Values are presented for the 59 conformations of a 12 x 12 grid whose \(ab \text{ initio}\) energies are within 12 kcal / mol of the lowest.

\(^c\)In units of kcal / mol.
TABLE XII.
The \((\text{coul}, E) - (\text{coul}, A)) - (\text{coul}, E) - (\text{coul}, A))\) Measure of Electrostatic Favoring of the Incumbent Conformation Obtained for Blocked Alanine Using a Collection of Multipole Electrostatic Models.\(^a\)\(^b\)

<table>
<thead>
<tr>
<th>Range</th>
<th>Moment-Moment(^c)</th>
<th>m-md</th>
<th>mdq-mdq</th>
<th>mdqo-mdqo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 — 5 and higher</td>
<td>4.0</td>
<td>7.1</td>
<td>8.7</td>
<td>9.0</td>
</tr>
<tr>
<td>1 — 4 and higher</td>
<td>4.1</td>
<td>9.0</td>
<td>9.4</td>
<td>8.9</td>
</tr>
<tr>
<td>1 — 3 and higher</td>
<td>4.1</td>
<td>6.8</td>
<td>9.2</td>
<td>9.8</td>
</tr>
<tr>
<td>1 — 2 and higher</td>
<td>4.1</td>
<td>6.3</td>
<td>8.8</td>
<td>18.6</td>
</tr>
</tbody>
</table>

\(^a\)In units of kcal/mol.
\(^b\)This collection of models allows resolution with respect to range and moment-moment contributions.
\(^c\)m, d, q, and o indicate monopole, dipole, quadrupole, and octopole.

a careful analysis of the conformational dependence of atomic multipoles shows many small changes to multipole components of all orders, all seeming to favor the incumbent conformation.

In addition to polarization, a second physical origin of flow of electron density with conformation is the quantum mechanical effect that also causes intrinsic torsional energy. For example, in ethane, as the conformation changes from staggered to eclipsed, .062 e of electron density flows out of the C—C bond, onto C, and into the C—H bonds. This flow is not polarization. For blocked alanine, changes in multipole components with conformation are largest for atoms forming a bond about which torsion angle rotation occurs. This is especially true for higher order moments such as quadrupole and octopole.

DISTANCE BETWEEN MODEL AND \(ab\) INITIO ENERGY SURFACES

Let G be the set of grid points of a 72 \(\times\) 72 grid for which the \(ab\) initio energy is within 16 kcal/mol of the lowest. Let \(e_{r,0}\) and \(e_{i,0}\) be the molecular mechanics model and \(ab\) initio energies, respectively, at grid point \((i, j)\). Let \(d\) be the separation between grid points. A measure of distance between model and \(ab\) initio energy surfaces was defined to be

\[
\sum_{(i,j) \in G} \left( \frac{1}{dS_{r,j}} \right)^2 \left( e_{r,j} - e_{i,j} \right)^2 \right)^{1/2}
\]

where \(dS_{r,j}\) is an estimate of the surface area of element of the \(ab\) initio energy surface that is associated with grid point \((i, j)\). \(dS_{r,j}\) was defined to be the sum of the surface areas of eight triangular tiles, formed from the points

\[
\begin{align*}
\delta & - \frac{\delta}{2}, \delta + \frac{\delta}{2}, \\
\delta & - \frac{\delta}{2}, \delta, \\
\delta & - \frac{\delta}{2}, - \frac{\delta}{2}, \\
\delta & - \frac{\delta}{2}, x, \\
\phi & - \frac{\delta}{2}, \phi, \\
\phi & - \frac{\delta}{2}, - \frac{\delta}{2}, \\
\phi & - \frac{\delta}{2}, x, \\
\phi & + \frac{\delta}{2}, \phi, \\
\phi & + \frac{\delta}{2}, - \frac{\delta}{2}, \\
\phi & + \frac{\delta}{2}, x
\end{align*}
\]

by connecting each of the eight outer points to the central point and to the two neighboring points. Less formally, eq. (20) is proportional to the root mean square (rms) over G of the distance between the surfaces along the normal to the \(ab\) initio surface. For each grid point \((i, j) \in G\), the ratio

\[
\frac{\left( \frac{1}{dS_{r,j}} \right)^2}{\sum_{(i,j) \in G} \frac{1}{dS_{r,j}}} \frac{1}{\left( e_{r,j} - e_{i,j} \right)^2}
\]

JOURNAL OF COMPUTATIONAL CHEMISTRY
which occurs in eq. (20) is dependent on the unit of circular arc that is used to measure torsion angles. Here, the unit of circular arc is chosen to be 30°.

This measure of distance is similar in spirit to a root mean square deviation (rmsd), the difference being that steeper regions of the \( \text{ab initio} \) energy surface are weighted less than flatter regions. As a result of this weighting, our distance between model and \( \text{ab initio} \) energy surfaces is, for all models, smaller than a straight rmsd. For most models, the rmsd between model and \( \text{ab initio} \) energy surfaces is dominated by the steepest parts of the \( \text{ab initio} \) energy surface, which cover only a small fraction of the area of the grid. For this reason, rmsd is not the most useful measure of distance between energy surfaces.

Results and Conclusions

In this section, relative accuracies are determined for our collection of electrostatic models by comparison to the target \( \text{ab initio} \) energy surface.

Of primary concern is the effect on accuracy of allowing nonspherical atomic charge densities.

For \((\alpha, \phi)\) of blocked alanine, the six electrostatic models of Table III are combined with the seven \([\text{rep}] + [\text{tor}]\) models of Table VI to give 42 molecular mechanics models. For each of these models, the corresponding \([\text{rep}] + [\text{tor}] + [\text{elec}]\) energy surface was restricted to a range of \([0.0, 80.0]\) kcal/mol. The zero of energy was chosen to be the global minimum over the \(72 \times 72\) grid on which the energy surfaces are defined, and energies greater than 80.0 kcal/mol were truncated to 80.0. The \( \text{ab initio} \) energy surface was similarly restricted to a range of \([0.0, 20.0]\) kcal/mol.

In Figure 9, \([\text{rep}] + [\text{tor}] + [\text{elec}]\) energy surfaces are plotted for the molecular mechanics models \([\text{merck}] + [\text{pd}]\) and \([\text{decomp}] + [\text{cpk}]\). A comparison of Figures 2 and 9a shows that, for standard representations of \([\text{elec}]\) and \([\text{rep}] + [\text{tor}]\), differences between model and \( \text{ab initio} \) energy surfaces can be large.

Table XIV shows the distance between model and \( \text{ab initio} \) energy surfaces for each of our 42 models. For most of the 12 standard models, the \([\text{mul}]\) and \([\text{pd}]\) electrostatic models combined with \([\text{rep}] + [\text{tor}]\) models other than \([\text{decomp}]\), the weighted average error is greater than 1.50 kcal/mol. Electrostatic energy differences between conformations would be reduced by (1) uniformly scaling atomic charge densities to account for electron correlation and (2) including the effects of a dielectric medium. Therefore, with proper care, errors introduced into applications can be reduced relative to those presented in Table XIV. However, none of the standard models is expected to have the accuracy needed to enable reliable structure prediction. Finally, from Table XIV, we see that the errors that remain in our standard models are large relative to the errors that are expected in our target \( \text{ab initio} \) energy surface.

Figure 10 is an integer map of the difference between the \([\text{merck}] + [\text{pd}]\) model and the \( \text{ab initio} \) energy surfaces. For each grid point of a \(24 \times 24\) grid, the actual error has been truncated to an integer. This pattern of error (too high in the E
TABLE XIV. Distances between Model and Ab initio Energy Surfaces

<table>
<thead>
<tr>
<th></th>
<th>[mul]</th>
<th>[pd]</th>
<th>[md]</th>
<th>[mdq]</th>
<th>[mdqo]</th>
<th>[cpk]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[opls]</td>
<td>3.111</td>
<td>2.125</td>
<td>1.788</td>
<td>1.702</td>
<td>2.630</td>
<td>1.682</td>
</tr>
<tr>
<td>[amber]</td>
<td>2.577</td>
<td>1.902</td>
<td>1.517</td>
<td>1.367</td>
<td>2.007</td>
<td>1.337</td>
</tr>
<tr>
<td>[eceppl]</td>
<td>2.618</td>
<td>1.840</td>
<td>1.349</td>
<td>1.328</td>
<td>2.139</td>
<td>1.117</td>
</tr>
<tr>
<td>[merck]</td>
<td>2.571</td>
<td>1.695</td>
<td>1.412</td>
<td>1.240</td>
<td>2.210</td>
<td>1.174</td>
</tr>
<tr>
<td>[decomp]</td>
<td>2.350</td>
<td>1.758</td>
<td>1.520</td>
<td>1.253</td>
<td>1.891</td>
<td>1.022</td>
</tr>
<tr>
<td>[mm3]</td>
<td>2.205</td>
<td>1.464</td>
<td>1.196</td>
<td>1.327</td>
<td>2.282</td>
<td>1.547</td>
</tr>
</tbody>
</table>

*aThe distance between model and ab initio energy surfaces was minimized by addition of a constant function (one independent parameter).

region, too low in the G region) is independent of the [rep] + [tor] model with which the [pd] electrostatic model is combined.

Of our 42 molecular mechanics models, the most accurate is [decomp] + [cpk], with a weighted average error of 1.022 kcal/mol. A rough ordering of the seven [rep] + [tor] models from softest to hardest is given by [mm3], [decomp], [merck], [eceppl], [amber], [opls], and [charmm]. For most of the six [elec] models, the fits between model and ab initio energy surfaces are best for the softest models [mm3], [decomp], and [merck] and worst for the hardest models [opls] and [charmm]. For the better [elec] models [mdq] and [cpk], the fits obtained with the softest [rep] + [tor] model, [mm3], are not as good as those obtained with the slightly harder models [decomp] and [merck]. This result, together with integer maps of the difference between model and ab initio energy surfaces, is consistent with the observation by Halgren et al. and by Waldman and Hagler 38 that an exponential representation of repulsion is probably too soft.

Of the two partial charge electrostatic models, [pd] is expected to be more accurate than [mul]. For all of the [rep] + [tor] models, [pd] does better than [mul]. This suggests that our measure of distance between surfaces is capable of determining relative accuracies for a collection of electrostatic models.

The three multipole electrostatic models relax the restriction on atomic charge densities of spherical symmetry but approximate 1–2 and 1–3 electrostatic interaction energies. For all but one of the [rep] + [tor] models, [md] and [mdq] do better than [pd]. This suggests that accurate representation of electrostatic energy requires a careful treatment of the nonspherical nature of atomic charge densities. For both [md] and [mdq], removing electrostatic interaction energies of types 1–2 and 1–3 causes distances between model and ab initio energy surfaces to increase. This suggests that accurate representation of electrostatic energy requires that contributions from 1–2 and 1–3 interactions be included. It also suggests that representation of 1–2 and 1–3 electrostatic interaction energies using lower order multipole–multipole interaction energies does less harm than good, despite errors introduced by lack of physical meaning.

The nonspherical CPK electrostatic model removes all restrictions on the shapes of atomic charge densities and calculates electrostatic interaction energies exactly. For all but one of the [rep] + [tor] models, [cpk] does better than [md] and [mdq]. The one exception to this rule is the
model [mm3] + [cpk], where the extreme softness of [mm3] in the region of C,„„, combines with the extreme stability of [cpk] to cause a low prediction in this region. Of our six electrostatic models, [cpk] is the most accurate. This result further supports our primary conclusion that accurate representation of electrostatic energy requires both careful treatment of the nonspherical nature of atomic charge densities and inclusion of 1—2 and 1—3 interactions.

For the 12 standard models, Table XIV implies an electrostatic origin for at least part of the difference between model and ab initio energy surfaces. Errors in a estimate of [ele] that depend on only one torsion angle degree of freedom can be corrected for by compensating errors in the usual one-dimensional Fourier series representation of intrinsic torsional energy. Of greater concern are errors that depend simultaneously on two torsion angle degrees of freedom. Given the range of functional forms that is used by the current generation of protein potential functions, these errors can not be corrected for by compensating for errors in the other components. Because 1—3 electrostatic interaction energies are predicted to be nonnegligible, accurate representation of electrostatic energy using a partial charge model in combination with the usual one-dimensional Fourier series representation of intrinsic torsional energies is predicted to require a two-dimensional correction. This correction is expected to have a natural dependence on pairs of adjacent torsion angles as opposed to interatomic distances.

To determine the nature of the difference between model and ab initio energy surfaces, this difference was corrected by addition of a sum of one-dimensional Fourier series of order 6:

\[
a_0 + \sum_{m=1}^{6} \left( a_m \cos(m\phi) + b_m \sin(m\phi) \right) + \sum_{m=1}^{6} \left( c_m \cos(m\psi) + d_m \sin(m\psi) \right)
\]

The 25 independent parameters of this one-dimensional correction were obtained by minimizing the distance between model and ab initio energy surfaces. A single local minimum was indicated by minimization from many different starting points. Table XV shows the distance between the corrected-model energy surface and the ab initio energy surface for each of our 42 models. For most of the corrected models, the weighted average error is about .5 kcal/mol. Clearly, a more accurate representation of the ab initio energy surface requires the addition of a nonnegligible two-dimensional correction—in other words, a function of (\phi, \psi) as opposed to a function of \phi plus a function of \psi. Also, none of the electrostatic models is consistently better than others. The expected progression in the accuracy with which two-dimensional electrostatic contributions are represented is not detectable. This suggests that the two-dimensional corrections, needed to make model energy surfaces match the ab initio energy surface, are dominated by something other than a poor representation of 1—3 electrostatic interaction energies. By default, the origin of this error must be in the representation of repulsion, intrinsic torsional energy, or polarization. The conclusion that two-dimensional corrections are dominated by errors introduced by our models of [rep] + [tor] is sup-

| TABLE XV. Distances between Corrected-Model and Ab Initio Energy Surfaces. |
|------------------|---|---|---|---|---|
|                  | [mul] | [pd] | [md] | [mdq] |
| charm]           | 0.731 | 0.792 | 0.777 | 0.746 | 0.769 | 0.704 |
| op]              | 0.460 | 0.526 | 0.515 | 0.495 | 0.481 | 0.525 |
| amber]           | 0.472 | 0.513 | 0.530 | 0.569 | 0.544 | 0.527 |
| ecepp]           | 0.467 | 0.499 | 0.513 | 0.578 | 0.567 | 0.621 |
| merck]           | 0.391 | 0.461 | 0.442 | 0.420 | 0.415 | 0.474 |
| decomp]          | 0.732 | 0.810 | 0.717 | 0.504 | 0.540 | 0.503 |
| mm3]             | 0.639 | 0.640 | 0.601 | 0.733 | 0.784 | 0.747 |

Model energy surfaces were corrected by addition of a sum of one-dimensional Fourier series of order 6 (25 independent parameters).
ported by the large differences between \( \text{rep} \) + \( \text{tor} \) models that are seen in Figures 3 and 7. The relative success of the crude \( \text{rep} \) + \( \text{tor} \) model \( \text{decomp} \), when combined with \( \text{cpk} \), suggests that if errors introduced by models of \( \text{rep} \) + \( \text{tor} \) could be reduced, then the expected variation in the accuracy with which our electrostatic models represent two-dimensional electrostatic contributions might become detectable.

To understand better the two-dimensional nature of the difference between model and \( \text{ab initio} \) energy surfaces, this difference was corrected by addition of a two-dimensional Fourier series of order 6:

\[
a_0 + \sum_{m=1}^{n} (a_m \cos(m\phi) + d_m \sin(m\phi)) + \sum_{n=1}^{N} (a_n \cos(n\phi) + d_n \sin(n\phi))
\]

\[
+ \sum_{m=1}^{n} \sum_{n=1}^{N} (a_{mn} \cos(m\phi) \cos(n\phi) + b_{mn} \cos(m\phi) \sin(n\phi) + c_{mn} \sin(m\phi) \cos(n\phi) + d_{mn} \sin(m\phi) \sin(n\phi))
\]

(24)

The 169 independent parameters of this two-dimensional correction were obtained by minimizing the distance between model and \( \text{ab initio} \) energy surfaces. Again, a single local minimum was indicated by minimization from different starting points. Table XVI shows the distance between corrected-model and \( \text{ab initio} \) energy surfaces for each of our 42 models. Here, the fits are almost perfect. Whatever the origin of the difference between model and \( \text{ab initio} \) energy surfaces, it is better fit by a function that depends simultaneously on both \( \phi \) and \( \psi \). Figure 11 is a contour plot of the corrected \( \text{merck} + \text{pd} \) energy surface.

For many of the models, the two-dimensional terms of the correction are much larger than would be expected for electrostatic or intrinsic torsional energies. These large two-dimensional terms are mainly compensating for errors in repulsion. Electrostatic interaction energies of type 1—3, which are neglected by partial charge models, are predicted to be nonnegligible by multipole models. To estimate the magnitudes of purely two-dimensional contributions to a typical 1—3 electrostatic interaction energy, a two-dimensional

| TABLE XVI. Distances between Corrected-Model and Ab Initio Energy Surfaces.6 |
|--------------|--------------|--------------|--------------|
| [md]         | [pd]         | [md]         | [mdq]        |
| [charm]      | 0.201        | 0.192        | 0.206        | 0.208        |
| [osp]        | 0.059        | 0.054        | 0.066        | 0.066        |
| [amb]        | 0.108        | 0.110        | 0.107        | 0.112        |
| [cccp]       | 0.132        | 0.122        | 0.113        | 0.130        |
| [merk]       | 0.028        | 0.034        | 0.031        | 0.034        |
| [decomp]     | 0.047        | 0.049        | 0.058        | 0.059        |
| [mm3]        | 0.199        | 0.196        | 0.205        | 0.208        |

6Model energy surfaces were corrected by addition of a two-dimensional Fourier series of order 6 (169 independent parameters).
Fourier series of order 3 was fit to the energy surface that results from the \((N_{33}, C_{05})\) multipole–multipole interaction in the model [mdqo]. The 49 independent parameters were obtained by minimizing the rmsd over a \(24 \times 24\) grid. The fit is exact because each moment–moment contribution through octopole can be represented exactly by a two-dimensional Fourier series of order 3. The magnitudes of the coefficients of two-dimensional terms are \(\sim 0.3\) kcal/mol. Each such coefficient would create energy differences between conformations of \(\sim 0.6\) kcal/mol. This truncated multipole expansion has not converged and may not be physically meaningful, but it provides an estimate of the purely two-dimensional contributions to 1–3 electrostatic interaction energies that are predicted by more accurate nonspherical CPK electrostatic models.

**Discussion**

This work, which has focused on the \([\text{elec}]\) component of the \((\phi, \psi)\) energy surface of blocked alanine, should be viewed in the context of a strategy for obtaining a complete protein potential function that is accurate enough to enable reliable structure prediction. In this work, we succeeded in using the \(ab\) \(\text{initio}\) energy surface for \((\phi, \psi)\) of blocked alanine (1) to construct electrostatic models for blocked alanine that are more accurate than standard partial charge models, and (2) to correct for some of the error in other energy components by introducing compensating errors into a two-dimensional Fourier series representation of intrinsic torsional energy. Ideally, a complete set of \(ab\) \(\text{initio}\) energy surfaces, one for each pair of adjacent torsion angles of each blocked amino acid, might be used (1) to construct accurate electrostatic models for each of the blocked amino acids, and (2) to parameterize a two-dimensional Fourier series representation of intrinsic torsional energy over the complete set of energy surfaces. A complete protein potential function that could represent with high accuracy the large collection of intramolecular energies that is equivalent to a complete set of \(ab\) \(\text{initio}\) energy surfaces might be accurate enough to enable reliable structure prediction. Any strategy for using a complete set of \(ab\) \(\text{initio}\) energy surfaces in the construction of an accurate molecular mechanics model is complicated by the following problems: (1) the possibility of nonnegligible \([\text{pol}]\), (2) elimination of polarization-induced deformations from atomic charge densities, and (3) accurate representation of \([\text{rep}]\). These problems are the subject of the remaining sections.

**NEGLECT OF POLARIZATION**

Our strategy for obtaining an accurate protein potential function is based on the following assumption. A functional form that includes (1) an accurate multipole representation of \([\text{elec}]\), (2) an accurate function of interatomic distances representation of \([\text{rep}]\), and (3) an accurate two-dimensional Fourier series representation of \([\text{tor}]\) but neglects (1) \([\text{pol}]\), except indirectly through the use of a dielectric continuum, and (2) anisotropic repulsion or dispersion can become accurate enough to enable reliable structure prediction. Implicit within this assumption is a more basic assumption concerning the ultimate limit to the accuracy of a nonpolarizable model. We recognize the existence of some limit, analogous to the Hartree-Fock limit in quantum chemistry, such that, within the restriction of a nonpolarizable model, a potential function cannot become more accurate than this limit. Our strategy assumes that (1) the current generation of nonpolarizable models is far from this limit, and (2) this limit will be accurate enough to enable reliable structure prediction.

**A SET OF IDEALIZED UNPOLARIZED ATOMIC CHARGE DENSITIES**

To clarify the following discussion, we will assume that the set of conformations over which an energy surface is calculated is obtained by varying two adjacent torsion angles over a \(24 \times 24\) grid. All other torsion angles are held fixed at one specific value. Calculation of energy surfaces over a grid of three or more dimensions is not currently feasible. As a consequence, for many of the blocked amino acids (the exceptions being alanine, glycine, and proline), the collection of conformations for which wave functions are available is concentrated within a small region of the space of conformations. For example, for blocked asparagine, there exist four significant pairs of adjacent torsion angles: \((\phi, \psi), (\phi, \chi_1), (\psi, \chi_1),\) and \((\chi_1, \chi_2)\). The set of conformations over which the \((\chi_1, \chi_2)\) energy surface is calculated has a single value of \((\phi, \psi)\), which we will refer to as \((\phi', \psi')\). As a consequence, of the wave functions that are available from the four energy surfaces of blocked as-
paragine, a disproportionate number have \((\phi, \psi) = (\phi', \psi')\).

For blocked alanine, in an attempt to obtain a set of idealized unpolarized atomic multipoles, 59 sets of atomic multipoles, each obtained from a different wave function, were averaged. If, for blocked asparagine, atomic multipoles were averaged over all available wave functions, then the resulting multipole electrostatic models would be expected to contain an inaccurate energetic preference for \((\phi, \psi) = (\phi', \psi')\). Furthermore, our results for blocked alanine concerning electrostatic favoring of the incumbent conformation indicate that multipole electrostatic models can become less accurate than even crude partial charge electrostatic models when polarization-induced deformations of atomic charge densities are not properly removed. This problem is equivalent to that which most severely limits the utility of existing libraries of atomic multipoles.\(^{39,40}\)

**REPULSION**

Our results indicate that, in addition to the [elec] component, possible sources of two-dimensional contributions to differences between model and \(ab\) initio energy surfaces include \(\text{[rep]}\), \(\text{[tor]}\), and \(\text{[pol]}\). This suggests a practical need for two-dimensional corrections to estimates of \(\text{[rep]}\) + \(\text{[tor]}\). The obvious choice, corrections in the form of a two-dimensional Fourier series, must be used carefully because of the possibility that two-dimensional contributions to error from \(\text{[rep]}\) might not be independent of a third adjacent torsion angle degree of freedom.

The \((\phi, \psi)\) energy surface of blocked alanine is a two-dimensional slice through a \((22 \times 3 - 6)\)-dimensional space of conformations. The corrected model function should reproduce \(ab\) initio energies on this two-dimensional slice. The corrected model function should also reproduce \(ab\) initio energies at low-energy points throughout the entire space. To accomplish this, the functional form that is used for the correction must depend on internal coordinates and interatomic distances in a physically reasonable way. For example, the \((O_{12}, H_{22})\) repulsion is dependent on torsion angles \(\omega, \phi\), and \(\chi_1\). Change to \(\omega\) or \(\chi_1\) might greatly reduce a repulsion that depends on distance as the inverse twelfth power. A two-dimensional Fourier series that compensates on the \((\phi, \psi)\) energy surface for errors in an estimate of this repulsion would cause large errors between corrected model and \(ab\) initio energies at low-energy points off the slice. This situation cannot be corrected by addition of two-dimensional Fourier series with respect to \((\omega, \phi)\) and \((\phi, \chi_1)\). In general, any contribution to \(\text{[rep]}\) that depends simultaneously on three or more adjacent torsion angles will not be accurately represented by a sum (over pairs of adjacent torsion angles) of two-dimensional Fourier series.

**Acknowledgment**

We thank J. A. Grant for several helpful discussions. This work was supported by N. I. H. Program Project Grant P01 GM24483. Source code for computer programs developed by the authors as a part of this work is available electronically on the Internet via anonymous ftp from dasher.wustl.edu (128.252.162.151). Also available as postscript files from dasher.wustl.edu are Figures 1 to 11 and supplementary material consisting of 12 additional \((\phi, \psi)\) energy-surface contour plots.

**References**


22. SPARTAN version 2.0, Wavefunction, Inc., 18401 Von Karman, #210, Irvine, CA 92715.


