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 (ϕ, ψ) 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)

*Author to whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 16, No. 7, 791–816 (1995)

CCC 0192-8651 / 95 / 070791-26

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 *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.

Introduction

In this article we look at the effect on the (ϕ, ψ) energy surface of blocked alanine of substituting increasingly accurate representations of the electrostatic energy. More specifically, we (1) calculate a target *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 *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 *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.[1-7]

For small nonpolar organic molecules, the best current molecular mechanics force fields allow calculation of structure and energetics to within experimental accuracy.[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.[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.[9] In this article, we focus on the vacuum potential energy. Within the framework of molecular mechanics,[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.[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.[17] A method for obtaining atomic multipoles from wave functions has been suggested by Stone.[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 intermolecular electrostatic interaction energies.[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 [*qm*] 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

$$[qm] = [kin] + [coul] + [exch] \quad (1)$$

Here [*kin*] is defined to include the kinetic energy integrals, and [*coul*] is defined to include nuclear–nuclear repulsion; the electron–nuclear attraction integrals; and the electron–electron repulsion Coulomb integrals, including

$$\frac{1}{2}\sum_{j\in\{1,\ldots,N\}}\int \psi_j^*(\mathbf{x}_1)\psi_j(\mathbf{x}_1)\frac{1}{|\mathbf{r}_2-\mathbf{r}_1|} \times \psi_j^*(\mathbf{x}_2)\psi_j(\mathbf{x}_2)\,d\mathbf{x}_1\,d\mathbf{x}_2 \quad (2)$$

where ψ_j is the jth molecular spin orbital, $\mathbf{x}_1$ the space-spin variable of electron 1, $\mathbf{r}_1$ the space component of $\mathbf{x}_1$, and N the number of electrons. Moreover, [*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 [*coul*] causes [*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 ([*coul*]) and quantum effects ([*kin*] + [*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

$$[qm] = [elec] + [pol] + [rep] + [tor] + [geom] \quad (3)$$

where [*elec*] is the electrostatic energy, [*pol*] the polarization energy, [*rep*] the sum of repulsion and dispersion energies, [*tor*] the intrinsic torsional energy, and [*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 [*qm*] 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 [*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 [*coul*] component of an *ab initio* energy surface to construct reasonable estimates of [*elec*], [*p*], and [*r*]—the [*coul*] subcomponents of [*elec*], [*pol*], and [*rep*] + [*tor*], respectively. This decomposes [*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.

	[*elec*]	[*pol*]	[*rep*] + [*tor*]
[*kin*]	—	[τ_1]	[τ_2]
[*exch*]	—	[ϵ_1]	[ϵ_2]
[*coul*]	[*elec*]	[*p*]	[*r*]

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 **C** be the set of conformations over which an energy surface is defined. For each conformation $x \in \mathbf{C}$, (1) charge density is taken from the wave function Ψ_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 $y \in \mathbf{C}$, atomic charge densities are translated and rotated for consistency with nuclear coordinates; and (4) $[\mathbf{coul}]_x(y)$ is defined to be the electrostatic energy of the resulting molecular charge density. **[coul]** is an extension of $[coul]$ in the sense that

$$[coul](y) = [\mathbf{coul}]_y(y) \tag{4}$$

An estimate of $[elec]$ is obtained using

$$[elec](y) = \left(\frac{1}{\#\mathbf{C}}\right) \sum_{x \in \mathbf{C}} [\mathbf{coul}]_x(y) \tag{5}$$

where **#C** is the number of conformations in **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

$$[\mathbf{coul}]_x(y) = [elec](y) + [\mathbf{p}]_x(y) + [r]_x \tag{6}$$

[p] is an extension of $[p]$ in the sense that

$$[p](y) = [\mathbf{p}]_y(y) \tag{7}$$

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

$$\sum_{\substack{y \in \mathbf{C} \\ y \neq x}} ([\mathbf{coul}]_x(y) - ([elec](y) + [r]_x))^2 \tag{8}$$

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 $[\tau_1] + [\epsilon_1]$, so that

$$[p] + [\tau_1] + [\epsilon_1] \ll [p] \tag{9}$$

In other words, the $[coul]$ subcomponent of polarization is largely balanced by the $[kin] + [exch]$ subcomponent. This allows us to obtain reasonable representations of $[qm]$ 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 (ϕ, ψ) energy surface of blocked alanine (*N*-acetyl–N′-methyl alanineamide) was calculated at the HF/6-31G* level over a 24 × 24 grid using the *ab initio* molecular orbital program SPARTAN.[22] All bond lengths, bond angles, and torsion angles other than ϕ and ψ 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 II. The bond lengths and bond angles of the peptide group are those suggested by Benedetti.[23] Other elements of the geometry were taken from Momany et al.[1]

In the region of the C_{7axial} 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 × 72 grid, and our measure of distance between

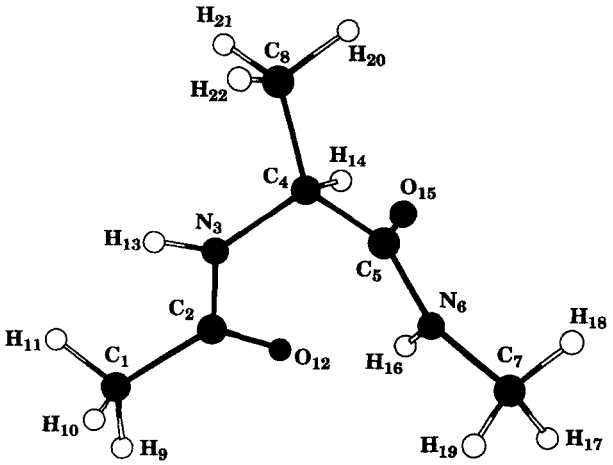

FIGURE 1. Notation used to identify atoms of blocked alanine.

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 ϕ or ψ.

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

$$\sum_{(i,j)\in \mathbf{H}} \left\{ \frac{(f_{(i+1,j)} - 2f_{(i,j)} + f_{(i-1,j)})(f_{(i+1,j)} - f_{(i,j)})}{\left[1 + (f_{(i+1,j)} - f_{(i,j)})^2\right]^{1/2}} + \frac{(f_{(i+1,j)} - 2f_{(i,j)} + f_{(i-1,j)})(f_{(i-1,j)} - f_{(i,j)})}{\left[1 + (f_{(i-1,j)} - f_{(i,j)})^2\right]^{1/2}} + \frac{(f_{(i,j+1)} - 2f_{(i,j)} + f_{(i,j-1)})(f_{(i,j+1)} - f_{(i,j)})}{\left[1 + (f_{(i,j+1)} - f_{(i,j)})^2\right]^{1/2}} + \frac{(f_{(i,j+1)} - 2f_{(i,j)} + f_{(i,j-1)})(f_{(i,j-1)} - f_{(i,j)})}{\left[1 + (f_{(i,j-1)} - f_{(i,j)})^2\right]^{1/2}} \right\} \quad (10)$$

Equation (10) was obtained as follows. Within a localized region of the (ϕ, ψ) 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 ψ can be approximated by the quadratic polynomial

$$F(\phi) = \frac{a}{2}(\phi - \phi_i)^2 + b(\phi - \phi_i) + c \quad (11)$$

TABLE II.
Geometry of Blocked Alanine in Z-matrix Format.

Bond lengths (Å)		Bond angles (degrees)		Torsion angles (degrees)	
C_2—C_1	1.510				
N_3—C_2	1.335	N_3—C_2—C_1	116.70		
C_4—N_3	1.449	C_4—N_3—C_2	121.90	C_4—N_3—C_2—C_1	180.00
C_5—C_4	1.522	C_5—C_4—N_3	109.30	C_5—C_4—N_3—C_2	ϕ
N_6—C_5	1.335	N_6—C_5—C_4	116.70	N_6—C_5—C_4—N_3	ψ
C_7—N_6	1.449	C_7—N_6—C_5	121.90	C_7—N_6—C_5—C_4	180.00
C_8—C_4	1.530	C_8—C_4—N_3	109.50	C_8—C_4—N_3—C_5	−123.53
H_9—C_1	1.090	H_9—C_1—C_2	109.47	H_9—C_1—C_2—N_3	180.00
H_{10}—C_1	1.090	H_{10}—C_1—C_2	109.47	H_{10}—C_1—C_2—N_3	−60.00
H_{11}—C_1	1.090	H_{11}—C_1—C_2	109.47	H_{11}—C_1—C_2—N_3	60.00
O_{12}—C_2	1.229	O_{12}—C_2—C_1	120.40	O_{12}—C_2—C_1—N_3	180.00
H_{13}—N_3	1.000	H_{13}—N_3—C_2	119.70	H_{13}—N_3—C_2—C_4	180.00
H_{14}—C_4	1.090	H_{14}—C_4—N_3	107.73	H_{14}—C_4—N_3—C_5	118.32
O_{15}—C_5	1.229	O_{15}—C_5—C_4	120.40	O_{15}—C_5—C_4—N_6	180.00
H_{16}—N_6	1.000	H_{16}—N_6—C_5	119.70	H_{16}—N_6—C_5—C_7	180.00
H_{17}—C_7	1.090	H_{17}—C_7—N_6	109.47	H_{17}—C_7—N_6—C_5	180.00
H_{18}—C_7	1.090	H_{18}—C_7—N_6	109.47	H_{18}—C_7—N_6—C_5	−60.00
H_{19}—C_7	1.090	H_{19}—C_7—N_6	109.47	H_{19}—C_7—N_6—C_5	60.00
H_{20}—C_8	1.090	H_{20}—C_8—C_4	109.47	H_{20}—C_8—C_4—N_3	180.00
H_{21}—C_8	1.090	H_{21}—C_8—C_4	109.47	H_{21}—C_8—C_4—N_3	−60.00
H_{22}—C_8	1.090	H_{22}—C_8—C_4	109.47	H_{22}—C_8—C_4—N_3	60.00

where

$$a = (f_{(i+1,j)} - 2f_{(i,j)} + f_{(i-1,j)}) \quad (12)$$

$$b = \left(\frac{f_{(i+1,j)} - f_{(i-1,j)}}{2}\right) \quad (13)$$

$$c = f_{(i,j)} \quad (14)$$

and where the unit of circular arc that is used to measure ϕ is taken to be the 5° separation between grid points. Over the segment of this curve that is defined by the domain $[\phi_i - \frac{1}{2}, \phi_i + \frac{1}{2}]$, the line integral of the square of the curvature is

$$\int_{[\phi_i - \frac{1}{2}, \phi_i + \frac{1}{2}]} \frac{a^2}{\left[1 + (a(\phi - \phi_i) + b)^2\right]^{3/2}} d\phi = \frac{a\left(\frac{a}{2} + b\right)}{\left[1 + \left(\frac{a}{2} + b\right)^2\right]^{1/2}} + \frac{a\left(\frac{a}{2} - b\right)}{\left[1 + \left(\frac{a}{2} - b\right)^2\right]^{1/2}} \quad (15)$$

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 surfce 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 × 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 (ϕ, ψ) 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.[24,25] 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.[26] 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.[24] 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.[25] 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

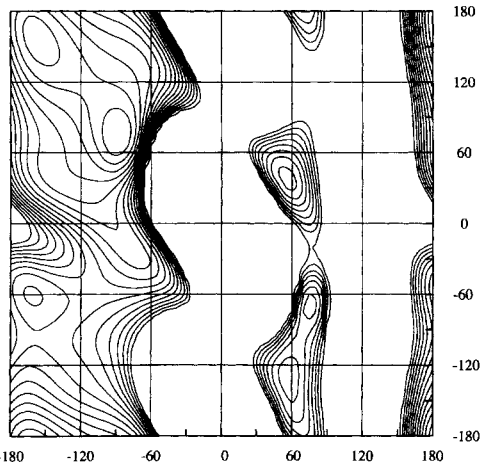

FIGURE 2. Contour plot of our target *ab initio* energy surface for (ϕ, ψ) of blocked alanine. The contour levels range from 1 to 16 kcal / mol in increments of 1 kcal / mol. This energy surface, defined over a 72 × 72 grid, results from HF / 6-31G* level calculations over a smaller 24 × 24 grid followed by interpolation. Bond lengths, bond angles, and torsion angles other than ϕ and ψ were maintained rigid.

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 atomic 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 interacion energy between N_{03} and C_{05} should change as a function of both ϕ and ψ. For models **[md]**, **[mdq]**, **[mdqo]**, and **[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 III.
Electrostatic Models.

Model	Type	Description
[mul]	Partial charge	Mulliken partial charges at nuclei
[pd]		Potential derived partial charges at nuclei
[md]	Multipole[a]	A monopole and dpole are included at each nucleus
[mdq]		A monopole, dipole, and quadrupole are included at each nucleus
[mdqo]		A monopole, dipole, quadrupole, and octopole are included at each nucleus
[cpk]	Nonspherical CPK[a]	Generalized atomic charge densities represented without approximation

[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 × 24 wave functions whose energies define our target *ab initio* energy surface for (ϕ, ψ) of blocked alanine, two sets of partial charges were calculated: Mulliken[27] and potential derived.[28] In an attempt to remove polarization contributions, both sets of partial charges were averaged over the 59 conformations of a 12 × 12 grid whose *ab initio* energies were within 12 kcal/mol of the lowest. Here, it was decided that a Boltzman weighted average would overemphasize conformations that contain favorable intramolecular interactions, because intermolecular interactions are not accounted for in the *ab initio* energies. The electrostatic energy surfaces of the models [**mul**] and [**pd**] were calculated using these averaged sets of partial charges over a 72 × 72 grid. For the 59 low-energy conformations that were used in averaging, changes with conformation are larger by about a factor of 3 for potential derived partial charges than for Mulliken partial charges.

The electrostatic energy surfaces of the multipole 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. Distributed atomic multipoles[18] were calculated from the wave functions whose energies define our target *ab initio* energy surface for (ϕ, ψ) 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 positions of the nucleus and its connected nuclei.[29] To remove polarization contributions, atomic multipoles were averaged over the same set of 59 conformations 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 × 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 interaction energy was approximated using a truncated multipole expansion.[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 developed by the authors.

The product of two Gaussians is itself a Gaussian, 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 multipoles from wave functions[18] consists of the following steps: (1) Multipoles are calculated at GG sites for elements of electron density corresponding to products of Gaussian primitives, and (2) multipoles 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 moments used a different origin.[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 electrostatic energy of a molecule that contains torsion angle degrees of freedom—differs slightly from the goal of multipole expansion in most previous work.[15] Consequently, the primary consideration in the assignment of elements of electron density to nuclei must be the movement with conformation that follows from this assignment. For this reason, our algorithm for the assignment of elements 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 approximation from the corresponding HF/6-31G* wave function. These sets of atomic charge densities were used to calculate 59 electrostatic energy surfaces over a 24 × 24 grid. For each set of atomic charge densities and for each conformation of the 24 × 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].

Range of parent nuclei	Assigned nucleus
1—1	Identical parent nuclei
1—2	Nearest nucleus If a GG site is equidistant from two (or more) nuclei, then equal contributions are moved to both (each)
1—3	Connecting nucleus
1—4 and higher	Nearest nucleus If a GG site is equidistant from two (or more) nuclei, then equal contributions are moved to both (each)

[a] The center of a product of two Gaussian primitives.
[b] Multipoles are calculated at GG sites for elements of electron density corresponding to products of Gaussian primatives and then "moved"[17] from GG sites to the assigned nuclei.
[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 × 24 grid, to a 72 × 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 ϕ or ψ, 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, 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 engry 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 initio* energies, this model is

TABLE V.
Assignment of GG Sites[a] to Nuclei for Electrostatic Model [cpk].[b]

Range of parent nuclei	Can the two parent nuclei of a GG site change relative position or orientation as a function ϕ or ψ?	
	No	Yes
1 — 1	Identical parent nuclei	
1 — 2	Closest parent nucleus	Closest (normalized distance)[c] parent nucleus If equidistant from the two parent nuclei, then equal contributions are moved to both
1 — 3	Connecting nucleus	Connecting nucleus
1 — 4 and higher	Closest nucleus within the rigid fragment containing the two parent nuclei	Closest nucleus within the rigid fragment containing the closest (normalized distance)[c] parent nucleus If equidistant from the two parent nuclei, then equal contributions are moved to both fragments

[a]The center of a product of two Gaussian primatives.
[b]A set of atomic charge densities is obtained from the molecular electron density by assigning each element of electron density to a nucleus.
[c]The closest parent nucleus is decided by comparing the normalized distances r_{1a}/R_a and r_{1b}/R_b, where r_{1a} ia the distance between the GG site and atom a and R_{aa} is the van der Waals radius of atom a.

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 *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 [*rep*] + [*tor*]. The 7 [*rep*] +

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

Model	Description
[amber]	Full atom amber[2,26]
[ecepp]	Modified ecepp[1,33] The large 12 -10 contributions to H bonds were removed
[opls]	opls[4]
[charmm]	charmm version 22[3]
[mm3]	mm3[5,8,34,35]
[merck]	merck molecular force field[6]
[decomp]	Estimate of repulsion + intrinsic torsional energy obtained by recognizing patterns of change in the quantum mechanics components [*kin*], [*coul*], and [*exch*] that are characteristic of repulsion and intrinsic torsional energy

[*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**], [**ecepp**], [**opls**], [**charmm**], [**mm3**], and [**merck**] were calculated over a 72 × 72 grid in the torsion angles ϕ and ψ. The model [**ecepp**] is essentially ecepp repulsion + dispersion.[33] 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

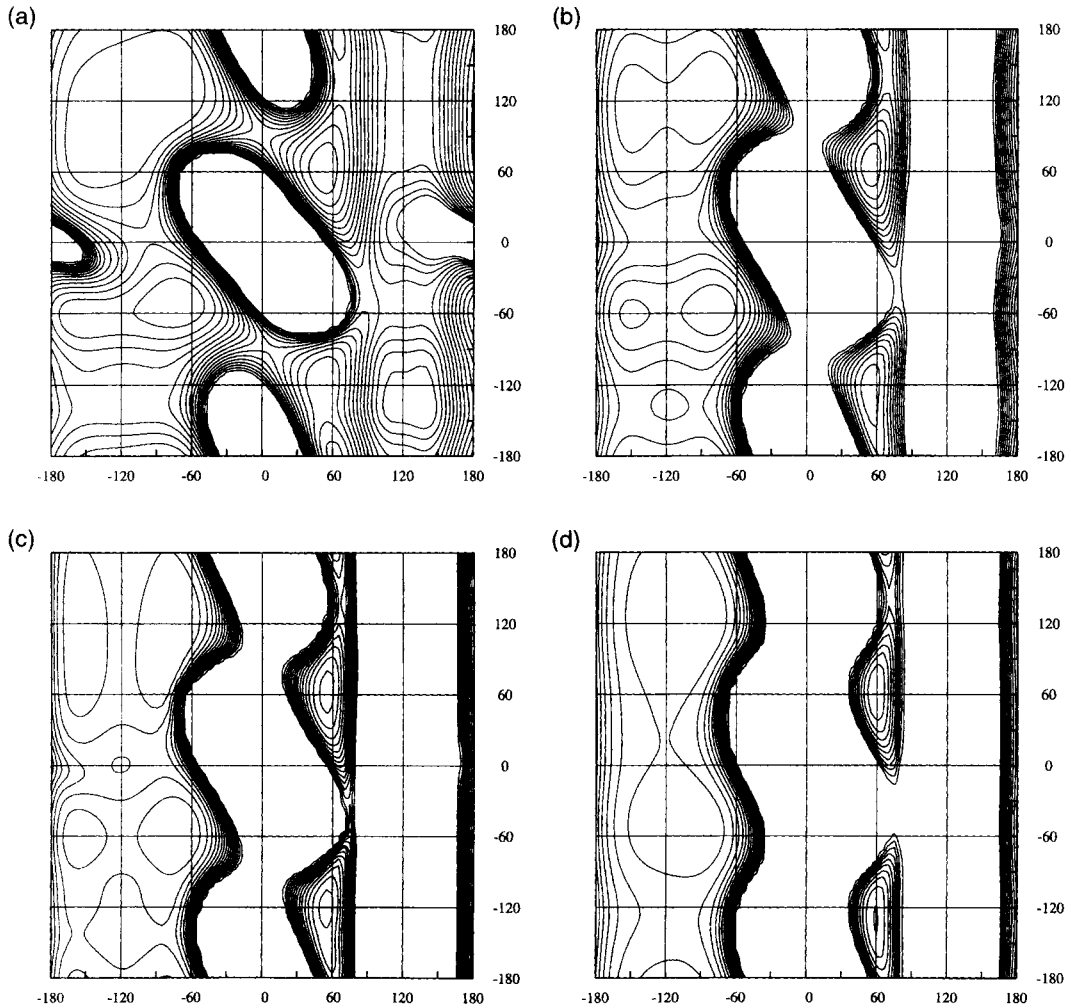

FIGURE 3. Contour plots of [*rep*] + [*tor*] energy surfaces for (ϕ, ψ) of blocked alanine of the models (a) [**mm3**], (b) [**merck**], (c) [**amber**], and (d) [**opls**]. The contour levels range from 1 to 16 kcal / mol in increments of 1 kcal / mol. Bond lengths, bond angles, and torsion angles other than ϕ and ψ were maintained rigid.

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 [*kin*], [*coul*], and [*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 χ. The *ab initio* energy curve was calculated over these 19 conformations at the HF/6-31G* level. Components [*kin*], [*coul*], and [*exch*] were obtained using programs developed by the authors. The corresponding pattern of change in the quantum mechanics components [*kin*], [*coul*], and [*exch*], along with the *ab initio* energy curve and the electrostatic energy curve of the model [**cpk**], are plotted in Figure 4.

As χ changes from 60° to 0°, both [*coul*] and [*exch*] decrease. These decreases are offset by a sharp increase in [*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 ([*coul*] + [*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

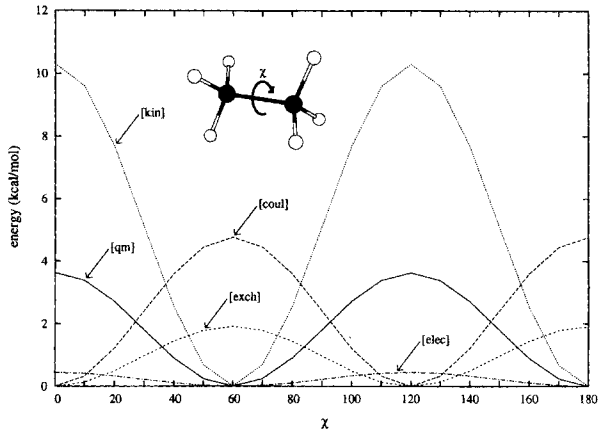

FIGURE 4. Pattern of change in the quantum mechanics components [*kin*], [*coul*], and [*exch*] corresponding to the change to the conformation of ethane that is specified in Table VII. Also included are the *ab initio* energy curve and the electrostatic energy curve of the model [**cpk**]. The *ab inition* energy curve was calculated at the HF / 6-31G* level. The electrostatic energy curve is the average of 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 $\chi = 110°$.

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 [*kin*], [*coul*], and [*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 Pitzer.[36]

As expected, the electrostatic energy curve of the model [**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

TABLE VII.
A Sequence of Conformations for Ethane (in Z-Matrix Format) over Which Changes in the Molecular Mechanics Components of the Energy Are Expected to be Concentrated in [*tor*].

Bond lengths (Å)		Bond angles (degrees)		Torsion angles (degrees)	
C_2—C_1	1.530				
H_3—C_1	1.090	H_3—C_1—C_2	109.47		
H_4—C_1	1.090	H_4—C_1—C_2	109.47	H_4—C_1—C_2—H_3	120.00
H_5—C_1	1.090	H_5—C_1—C_2	109.47	H_5—C_1—C_2—H_3	−120.00
H_6—C_2	1.090	H_6—C_2—C_1	109.47	H_6—C_2—C_1—H_3	χ[a]
H_7—C_2	1.090	H_7—C_2—C_1	109.47	H_7—C_2—C_1—H_6	120.00
H_8—C_2	1.090	H_8—C_2—C_1	109.47	H_8—C_2—C_1—H_6	−120.00

[a] $\chi \in \{0°, 10°, 20°, \ldots, 180°\}$.

densities that were obtained without approximation from the 12 (distinct) wave functions whose energies define the *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 *ab initio* energy curve. This value drops to ~ 3.0 kcal/mol with the inclusion of geometry optimization and electron correlation.[37] Whereas the [*qm*] curve is relatively insensitive to geometry optimization, the component curves [*kin*], [*coul*], and [*exch*] change considerably when bond lengths and bond angles are relaxed.[37]

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 *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 [*kin*], [*coul*], and [*exch*], along with the *ab*

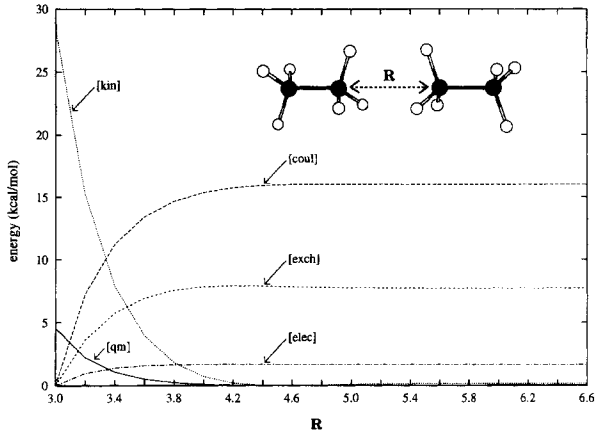

FIGURE 5. Pattern of change in the quantum mechanics components [*kin*], [*coul*], and [*exch*] corresponding to the change to the conformation of the ethane–ethane dimer that is specified in Table VIII. Also included are the *ab initio* energy curve and the electrostatic energy curve of the model [**cpk**]. The *ab initio* energy curve was calculated at the HF/631G* 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.

TABLE VIII.
A Sequence of Conformations for the Ethane – Ethane Dimer (in Z-Matrix Format) over Which Changes in the Molecular Mechanics Components of the Energy Are Expected to be Concentrated in [*rep*].

Bond lengths (Å)		Bond angles (degrees)		Torsion angles (degrees)	
C_2—C_1	1.530				
H_3—C_1	1.090	H_3—C_1—C_2	109.47		
H_4—C_1	1.090	H_4—C_1—C_2	109.47	H_4—C_1—C_2—H_3	120.00
H_5—C_1	1.090	H_5—C_1—C_2	109.47	H_5—C_1—C_2—H_3	−120.00
H_6—C_2	1.090	H_6—C_2—C_1	109.47	H_6—C_2—C_1—H_3	60.00
H_7—C_2	1.090	H_7—C_2—C_1	109.47	H_7—C_2—C_1—H_6	120.00
H_8—C_2	1.090	H_8—C_2—C_1	109.47	H_8—C_2—C_1—H_6	−120.00
C_9—C_1	R[a]	C_9—C_1—C_2	180.00	C_9—C_1—C_2—H_3	0.00
C_{10}—C_9	1.530	C_{10}—C_9—C_1	180.00	C_{10}—C_9—C_1—H_3	0.00
H_{11}—C_9	1.090	H_{11}—C_9—C_{10}	109.47	H_{11}—C_9—C_{10}—H_3	0.00
H_{12}—C_9	1.090	H_{12}—C_9—C_{10}	109.47	H_{12}—C_9—C_{10}—H_{11}	120.00
H_{13}—C_9	1.090	H_{13}—C_9—C_{10}	109.47	H_{13}—C_9—C_{10}—H_{11}	−120.00
H_{14}—C_{10}	1.090	H_{14}—C_{10}—C_9	109.47	H_{14}—C_{10}—C_9—H_{11}	60.00
H_{15}—C_{10}	1.090	H_{15}—C_{10}—C_9	109.47	H_{15}—C_{10}—C_9—H_{14}	120.00
H_{16}—C_{10}	1.090	H_{16}—C_{10}—C_9	109.47	H_{16}—C_{10}—C_9—H_{14}	−120.00

[a] $R \in \{3.0\text{ Å}, 3.2\text{ Å}, 3.4\text{ Å}, \ldots, 6.6\text{ Å}\}$.

initio energy curve and the electrostatic energy curve of the model ⟦**cpk**⟧, are plotted in Figure 5.

As R decreases from 6.6 to 3.0 Å, both [*coul*] and [*exch*] decrease. These decreases are offset by a sharp increase in [*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 decrease 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 θ 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 [*kin*], [*coul*], and [*exch*], along with the *ab initio* energy curve and the electrostatic energy curve of the model ⟦**cpk**⟧, are plotted in Figure 6.

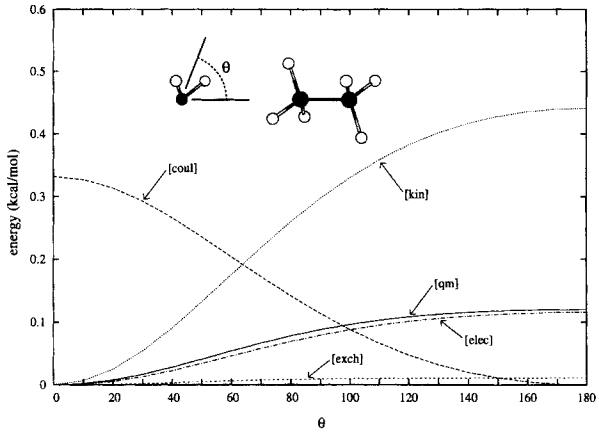

FIGURE 6. Pattern of change in the quantum mechanics components [*kin*], [*coul*], and [*exch*] corresponding to the change to the conformation of the water–ethane dimer that is specified in Table IX. Also included are the *ab initio* energy curve and the electrostatic energy curve of the model ⟦**cpk**⟧. The *ab initio* energy curve was calculated at the HF-6-31G* level. The electrostatic energy curve is the average of 19 curves, obtained using the 19 sets of atomic charge densities that were obtained without approximation from the HF/6-31G* wave functions of conformations with $\theta = 0°$ to $\theta = 180°$.

As θ changes from 0° to 180°, the decrease in [*coul*] is largely offset by an increase in [*kin*]. The [*exch*] curve is essentially flat. This compensation between the [*coul*] and [*kin*] components is consistent with the large [**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 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*].

Bond lengths (Å)		Bond angles (Å)		Torsion angles (degrees)	
C_2—C_1	1.530				
H_3—C_1	1.090	H_3—C_1—C_2	109.47		
H_4—C_1	1.090	H_4—C_1—C_2	109.47	H_4—C_1—C_2—H_3	120.00
H_5—C_1	1.090	H_5—C_1—C_2	109.47	H_5—C_1—C_2—H_3	−120.00
H_6—C_2	1.090	H_6—C_2—C_1	109.47	H_6—C_2—C_1—H_3	60.00
H_7—C_2	1.090	H_7—C_2—C_1	109.47	H_7—C_2—C_1—H_6	120.00
H_8—C_2	1.090	H_8—C_2—C_1	109.47	H_8—C_2—C_1—H_6	−120.00
O_9—C_1	6.00	O_9—C_1—C_2	180.00	O_9—C_1—C_2—H_3	0.00
X_{10}—O_9	1.00	X_{10}—O_9—C_1	θ[a]	X_{10}—O_9—C_1—H_3	0.00
H_{11}—O_9	0.98941	H_{11}—O_9—X_{10}	50.01345	H_{11}—O_9—X_{10}—C_1	90.00
H_{12}—O_9	0.98941	H_{12}—O_9—C_1	50.01345	H_{12}—O_9—C_1—C_1	−90.00

[a] $\theta \in \{0°, 10°, 20°, \ldots, 180°\}$.

$[pol] = [p] + [\tau_1] + [\epsilon_1]$ 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

$$[rep] + [tor] \approx (-1.68)([exch] - [exch]_{max}) \quad (16)$$

where $[exch]_{max}$ is the maximum of the $[exch]$ curve at $\chi = 60°$. For the ethane–ethane dimer, the ratio of change in repulsion to change in $[exch]$ is approximately $-.7$ when repulsion is small but increases slightly as repulsion increases. This relationship can be expressed more formally as

$$[rep] + [tor] \approx (-.700)([exch] - [exch]_{max}) + (.016)([exch] - [exch]_{max})^2 \quad (17)$$

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 ϕ and ψ, 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.

TABLE X.

Patterns of Change in Quantum Mechanics Components [*kin*], [*coul*], and [*exch*] Observed for Changes to the Conformations of Three Simplified Systems, Each Representative of an Isolated Increase in One of Three Molecular Mechanics Components—[*rep*], [*tor*], and [*pol*].

	[*kin*]	[*coul*]	[*exch*]
[*pol*]	↓ or ↑	↑ or ↓	
[*rep*]	↑	↓	↓
[*tor*]	↑	↓	↓

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

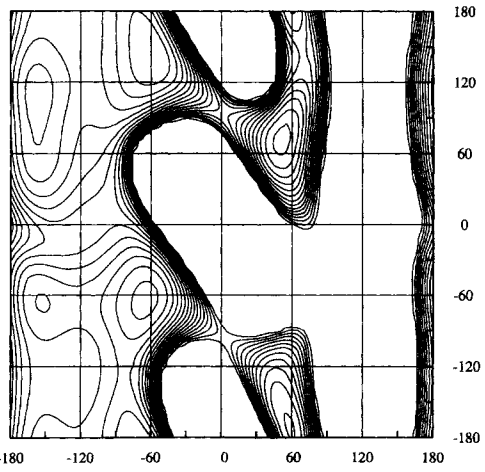

FIGURE 7. Contour plot of the [*rep*] + [*tor*] energy surface of the model [**decomp**] for (ϕ, ψ) of blocked alanine. The contour levels range from 1 to 16 kcal / mol in increments of 1 kcal / mol.

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 ρ_1 and ρ_2 from the positions of their nuclei $\mathbf{a}_1$ and $\mathbf{a}_2$. Derivation of the multipole-expansion representation of the electrostatic interaction energy requires that $r_1 + r_2 < |\mathbf{a}_2 - \mathbf{a}_1|$.[17,30-32] 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 ~ 1.5 Å) occur inside of the surfaces of the bonded atomic charge densities ($r_{\text{Oxygen}} \sim 1.60$ Å, $r_{\text{Nitrogen}} \sim 1.75$ Å, $r_{\text{Carbon}} \sim 1.90$ Å). Interactions of type 1—3 (distance ~ 2.5 Å) occur inside of the sum of the radii of the interacting atomic charge densities. Interactions of type 1—4 (distance ~ [2.9 Å, 3.8 Å]) 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 .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 .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, this 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 [**mdqo**] and [**mdq**], 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 $x \in \mathbf{C}$ of a polar molecule, the corresponding set of atomic charge densities $\{\rho_a : a \in \mathbf{A}\}$ is slightly deformed relative to the set of idealized unpolarized atomic charge densities $\{\overline{\rho_a} : a \in \mathbf{A}\}$. For each atom $a \in \mathbf{A}$ the small deformation $\delta\rho_a = (\rho_a - \overline{\rho_a})$ is a response to external electric fields created by $\{\rho_b : b \in \mathbf{A}, b \neq a\}$ in conformation x. The extended energy surface

$$[\mathbf{coul}]_x(y) = [elec](y) + [\mathbf{p}]_x(y) + [r]_x \quad (18)$$

is obtained by translation and rotation of $\{\rho_a : a \in \mathbf{A}\}$ with conformation according to the nonspherical CPK model. Here, $[\mathbf{p}]_x(y)$ is the contribution to energy differences between conformations that results from the polarization $\{\delta\rho_a : a \in \mathbf{A}\}$ of atomic charge densities. Without exception, $[\mathbf{p}]_x$ favors the incumbent conformation x.

For (ϕ, ψ) 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 $[\mathbf{coul}]_E(E)$ is lower than $[\mathbf{coul}]_E(A)$, that $[\mathbf{coul}]_A(A)$ is lower than $[\mathbf{coul}]_A(E)$, 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 $[\mathbf{p}]_x(x)$, the $[coul]$ component of $[pol]$. Table XII shows that, for the 59 conformations of a 12 × 12 grid whose *ab initio* energies are within 12 kcal/mol of the lowest, $[\mathbf{p}]_x(x)$ is large. Clearly, for any single conformation x, $[\mathbf{coul}]_x$ is a poor estimate of $[elec]$.

Because $[\mathbf{p}]_x$ 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 $[elec]$. For each conformation $x \in \mathbf{C}$, where **C** is the set of 59 conformations for which values of $[\mathbf{p}]_x(x)$ are presented in Table XII, the ratio of the average absolute value of $[\mathbf{p}]_x$

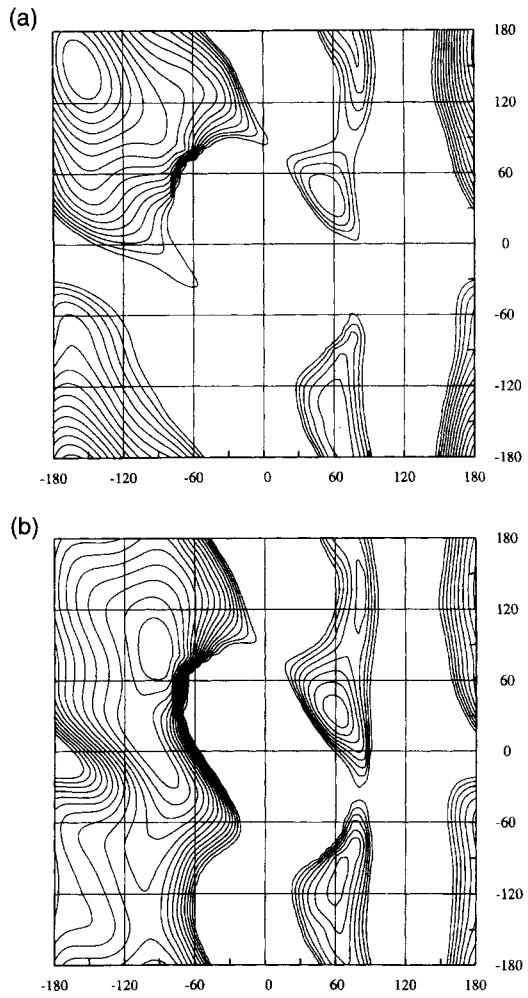

FIGURE 8. Electrostatic favoring of the incumbent conformation. Contour plots of model energy surfaces for (ϕ, ψ) of blocked alanine obtained by combining the $[rep] + [tor]$ energy surface of the model **[decomp]** with $[\mathbf{coul}]_{(\phi', \psi')}$ electrostatic energy surfaces calculated for 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.

defined by

$$\left(\frac{1}{\#\mathbf{C} - 1}\right) \sum_{\substack{y \in \mathbf{C} \\ y \neq x}} |[\mathbf{p}]_x(y)| \quad (19)$$

to $-[\mathbf{p}]_x(x)$ 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 $[\mathbf{p}]_x(y)$ lose coherence, tending to add less constructively, as y moves away from x. The shape

TABLE XI.
$[coul]_{(\phi_1,\psi_1)}(\phi_2,\psi_2)$ Obtained Using a Nonspherical CPK Electrostatic Model for Nine Low-Energy Conformations of Blocked Alanine.[a]

		(ϕ_1, ψ_1)								
(ϕ_2, ψ_2)		A	G	B	C	D	F	E	A*	C*
(−75, −45)	A	24.55	24.39	17.04	33.12	38.19	49.14	48.95	31.69	13.10
(−150, −55)	G	31.10	18.35	18.63	36.46	34.34	53.49	43.55	30.82	13.33
(−95, 10)	B	19.42	19.79	10.18	25.73	29.36	43.60	40.99	23.58	7.40
(−85, 75)	C	17.24	17.15	6.17	16.67	22.94	35.53	34.39	17.89	2.05
(−150, 80)	D	27.61	18.60	14.72	26.95	25.05	46.56	36.91	22.98	8.49
(−80, 150)	F	25.54	24.50	16.45	27.87	33.82	40.71	40.00	27.06	9.73
(−150, 150)	E	26.96	16.16	13.90	26.77	25.44	41.47	30.75	22.25	5.41
(60, 45)	A*	29.45	23.56	15.98	29.26	30.45	47.32	41.99	19.47	8.74
(60, −100)	C*	23.01	18.00	12.07	24.86	27.12	40.84	35.04	18.86	0.00

[a] In units of kcal / mol.

that is observed for $[\mathbf{p}]_x$ supports the expectation that error introduced by polarization into an estimate of $[elec]$ should be reduced by averaging.

Table XII shows that changes with conformations in $[\mathbf{p}]_x(x)$ 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 $[pol]$. 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 $[pol]$ with conformation will be small in comparison to the errors remaining in standard models of $[elec]$ and $[rep] + [tor]$. In other words, we expect that substituting increasingly accurate representations of electrostatic energy will reduce the distance between model and *ab 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 $([\mathbf{coul}]_A(E) - [\mathbf{coul}]_A(A)) - ([\mathbf{coul}]_E(E) - [\mathbf{coul}]_E(A))$ measure of electrostatic favoring of the incumbent conformation with respect to range and moment–moment contributions. The results, presented in Table XIII, indicate that $[\mathbf{p}]_x$ 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.
$[p]_{(\phi,\psi)}(\phi,\psi)$ Obtained for Blocked Alanine Using a Nonspherical CPK Electrostatic Model.[a,b,c]

	ϕ											
ψ	−180	−150	−120	−90	−60	−30	0	30	60	90	120	150
150	−6.8	−4.5	−3.2	−4.66	−7.4							
120	−6.4	−4.2	−2.7	−3.9	−7.5	−8.7						
90	−6.0	−4.1	−3.2	−4.5	−6.0							
60	−6.2	−4.7	−4.1	−6.1					−9.0			
30	−6.5	−5.2	−3.9	−5.1					−7.3			
0		−5.1	−4.1	−4.3								
−30	−6.9	−6.2	−5.5	−5.6	−7.2							
−60	−6.0	−5.4	−5.3	−6.0	−7.9							
−90	−5.8	−4.8	−4.7	−5.8					−3.8			
−120	−6.4	−4.7	−4.5	−5.4					−6.4			
−150	−7.0	−4.9	−4.3	−5.6					−8.0			
−180	−7.0	−4.9	−3.8	−5.3	−8.0							

[a] $[\mathbf{p}]_{(\phi,\psi)}(\phi,\psi)$ 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 × 12 grid whose *ab initio* energies are within 12 kcal / mol of the lowest.
[c] In units of kcal / mol.

TABLE XIII.
The ($[coul]_A(E) - [coul]_A(A)$) − ($[coul]_E(E) - [coul]_E(A)$) Measure of Electrostatic Favoring of the Incumbent Conformation Obtained for Blocked Alanine Using a Collection of Multipole Electrostatic Models.[a,b]

	Interactions included			
	Moment-Moment[c]			
Range	m — m	md-md	mdq-mdq	mdqo-mdqo
1 — 5 and higher	4.0	7.1	8.7	9.0
1 — 4 and hgher	4.1	9.0	9.4	8.9
1 — 3 and higher	4.1	6.8	8.2	9.8
1 — 2 and higher	4.1	6.3	8.8	18.6

[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 × 72 grid for which the *ab initio* energy is within 16 kcal/mol of the lowest. Let $e_{(i,j)}$ and $t_{(i,j)}$ be the molecular mechanics model and *ab initio* energies, respectively, at grid point (i, j). Let δ be the separation between grid points. A measure of distance between model and *ab initio* energy surfaces was defined to be

$$\left[\sum_{(i,j)\in\mathbf{G}}\left(\frac{\left(\frac{1}{dS_{(i,j)}}\right)^2}{\sum_{(k,l)\in\mathbf{G}}\left(\frac{1}{dS_{(k,l)}}\right)^2}\right)(t_{(i,j)}-e_{(i,j)})^2\right]^{1/2} \tag{20}$$

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

$$\begin{pmatrix}
\phi_i-\frac{\delta}{2}, & \psi_i+\frac{\delta}{2}, & \frac{t_{(i-1,j+1)}+t_{(i-1,j)}+t_{(i,j+1)}+t_{(i,j)}}{4}\\
\phi_i-\frac{\delta}{2}, & \psi_i, & \frac{t_{(i-1,j)}+t_{(i,j)}}{2}\\
\phi_i-\frac{\delta}{2}, & \psi_i-\frac{\delta}{2}, & \frac{t_{(i-1,j-1)}+t_{(i-1,j)}+t_{(i,j-1)}+t_{(i,j)}}{4}\\
\phi_i, & \psi_i+\frac{\delta}{2}, & \frac{t_{(i,j+1)}+t_{(i,j)}}{2}\\
\phi_i, & \psi_i, & t_{(i,j)}\\
\phi_i, & \psi_i-\frac{\delta}{2}, & \frac{t_{(i,j-1)}+t_{(i,j)}}{2}\\
\phi_i+\frac{\delta}{2}, & \psi_i+\frac{\delta}{2}, & \frac{t_{(i+1,j+1)}+t_{(i+1,j)}+t_{(i,j+1)}+t_{(i,j)}}{4}\\
\phi_i+\frac{\delta}{2}, & \psi_i, & \frac{t_{(i+1,j)}+t_{(i,j)}}{2}\\
\phi_i+\frac{\delta}{2}, & \psi_i-\frac{\delta}{2}, & \frac{t_{(i+1,j-1)}+t_{(i+1,j)}+t_{(i,j-1)}+t_{(i,j)}}{4}
\end{pmatrix} \tag{21}$$

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 $(i, j) \in \mathbf{G}$, the ratio

$$\left(\frac{\left(\frac{1}{dS_{(i,j)}}\right)^2}{\sum_{(k,l)\in\mathbf{G}} T\left(\frac{1}{dS_{(k,l)}}\right)^2}\right) \tag{22}$$

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 *ab initio* energy surface are weighted less than flatter regions. As a result of this weighting, our distance between model and *ab initio* energy surfaces is, for all models, smaller than a straight rmsd. For most models, the rmsd between model and *ab initio* energy surfaces is dominated by the steepest parts of the *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 *ab initio* energy surface. Of primary concern is the effect on accuracy of allowing nonspherical atomic charge densities.

For (ϕ, ψ) of blocked alanine, the six electrostatic models of Table III are combined with the seven [*rep*] + [*tor*] models of Table VI to give 42 molecular mechanics models. For each of these models, the corresponding [*rep*] + [*tor*] + [*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 × 72 grid on which the energy surfaces are defined, and energies greater than 80.0 kcal/mol were truncated to 80.0. The *ab initio* energy surface was similarly restricted to a range of [0.0, 20.0] kcal/mol.

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

Table XIV shows the distance between model and *ab initio* energy surfaces for each of our 42 models. For most of the 12 standard models, the ⟦**mul**⟧ and ⟦**pd**⟧ electrostatic models combined with [*rep*] + [*tor*] models other than ⟦**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 *ab initio* energy surface.

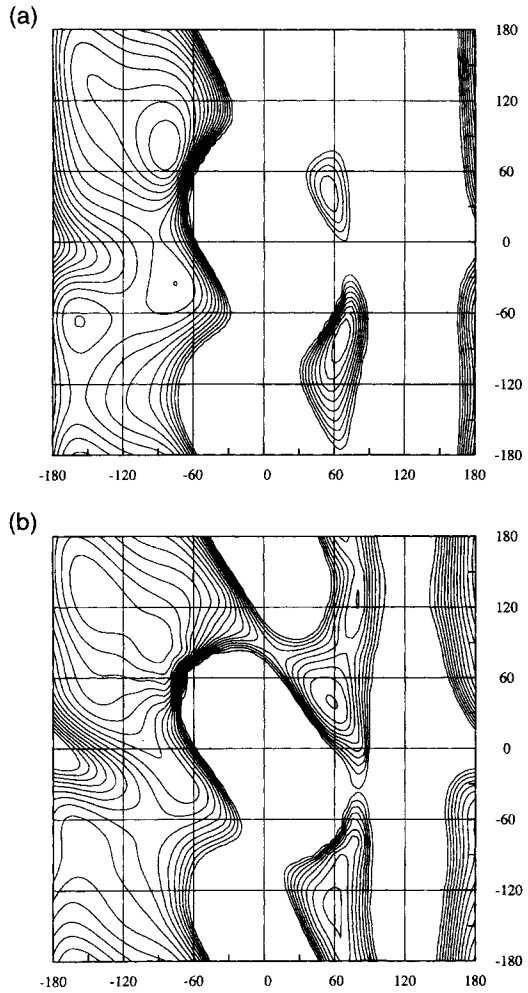

FIGURE 9. Contour plots of the [*rep*] + [*tor*] + [*elec*] energy surfaces of the models (a) ⟦**merck**⟧ + ⟦**pd**⟧ and (b) ⟦**decomp**⟧ + ⟦**cpk**⟧. The contour levels range from 1 to 16 kcal / mol in increments of 1 kcal / mol.

Figure 10 is an integer map of the difference between the ⟦**merck**⟧ + ⟦**pd**⟧ model and the *ab initio* energy surfaces. For each grid point of a 24 × 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.[a]

	[mul]	[pd]	[md]	[mdq]	[mdqo]	[cpk]
[charmm]	7.119	6.189	6.386	6.374	6.920	5.981
[opls]	3.111	2.125	1.788	1.702	2.630	1.662
[amber]	2.577	1.902	1.517	1.367	2.007	1.337
[ecepp]	2.618	1.840	1.349	1.328	2.139	1.117
[merck]	2.571	1.695	1.412	1.240	2.210	1.174
[decomp]	2.350	1.758	1.520	1.253	1.891	1.022
[mm3]	2.205	1.464	1.196	1.327	2.282	1.547

[a] The 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**], [**ecepp**], [**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[6] and by Waldman and Hagler[38] that an exponential representation of repulsion is probably too soft.

```
+4+3+2+1+1+1+1+1+1          +3+5          +7+4
+4+3+2+2+1+1 0+1+1            +5          +6+4
+3+2+2+1+1 0 0 0+1+2          +5          +5+3
+2+1+1+1 0 0 0 0 0+2+3                    +5+2
+1 0 0 0 0 0 0 0 0 0+3                    +4+1
+1 0 0 0 0 0 0-1-1-1+1                    +3+1
 0 0 0 0 0 0 0-1-3                        +2 0
 0 0 0 0 0 0 0-1-5          +1+1+3        +2 0
+1 0 0 0 0 0 0-1           +4+2+2+3       +3+1
+2+1+1+1 0 0 0-1            +3+3+3          +2
+1+1+1+1+1 0 0-1            +3+3+4          +1
    0 0+1+1+1 0-1             +2+4
     -1 0 0 0 0 0-1            +3
   -3-3-2 0 0 0 0 0             0
 0-2-3-2-2-1-1 0 0 0            0          +2 0
 0-1-2-2-2-2-2-1 0             -1          +2 0
 0-2-2-2-2-2-3-3-2-1 0          -1+1        +1 0
-1-2-2-2-2-2-3-3-3              -1+1         0-1
-1-2-2-2-2-2-3-3               -5-1          0-1
 0-1-1-1-1-1-2               -3-2 0         +1 0
 0 0 0 0 0 0-1              +1 0 0 0         +3 0
+2 0 0 0 0 0 0-1             +1+1+2         +4+2
+3+1 0 0 0 0 0 0             +2+2+3         +5+3
+4+2+1+1+1+1+1+1               +3+4         +6+4
+4+3+2+1+1+1+1+1+1             +3+5         +7+4
```

FIGURE 10. Integer map of differences between the [**merck**] + [**pd**] model and *ab initio* energy surfaces over the grid points of a 24 × 24 grid for which *ab initio* energy is within 16 kcal/mol of the lowest. The actual error is truncated to the corresponding integer.

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_{7axial} 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 an estimate of [*elec*] 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:

$$a_0 + \sum_{m=1}^{6} (a_m \cos(m\phi) + b_m \sin(m\phi)) + \sum_{m=1}^{6} (c_m \cos(m\psi) + d_m \sin(m\psi)) \quad (23)$$

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 (ϕ, ψ) as opposed to a function of ϕ plus a function of ψ. 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.[a]

	[**mul**]	[**pd**]	[**md**]	[**mdq**]		
[**charmm**]	0.731	0.792	0.777	0.746	0.769	0.704
[**opls**]	0.460	0.526	0.515	0.495	0.481	0.525
[**amber**]	0.472	0.513	0.530	0.569	0.544	0.627
[**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	0.764	0.747

[a]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 [*rep*] + [*tor*] models that are seen in Figures 3 and 7. The relative success of the crude [*rep*] + [*tor*] model [**decomp**] when combined with [**cpk**] suggests that if errors introduced by models of [*rep*] + [*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 *ab initio* energy surfaces, this difference was corrected by addition of a two-dimensional Fourier series of order 6:

$$\begin{aligned} a_{00} &+ \sum_{m=1}^{6} (a_{m0}\cos(m\phi) + d_{m0}\sin(m\phi)) \\ &+ \sum_{n=1}^{6} (a_{0n}\cos(n\psi) + d_{0n}\sin(n\psi)) \\ &+ \sum_{m=1}^{6}\sum_{n=1}^{6} (a_{mn}\cos(m\phi)\cos(n\psi) \\ &+ b_{mn}\cos(m\phi)\sin(n\psi) \\ &+ c_{mn}\sin(m\phi)\cos(n\psi) \\ &+ d_{mn}\sin(m\phi)\sin(n\psi)) \end{aligned} \quad (24)$$

The 169 independent parameters of this two-dimensional correction were obtained by minimizing the distance between model and *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 *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 *ab initio* energy surfaces, it is better fit by a function that depends simultaneously on both ϕ and ψ. Figure 11 is a contour plot of the corrected [**merck**] + [**pd**] energy surface.

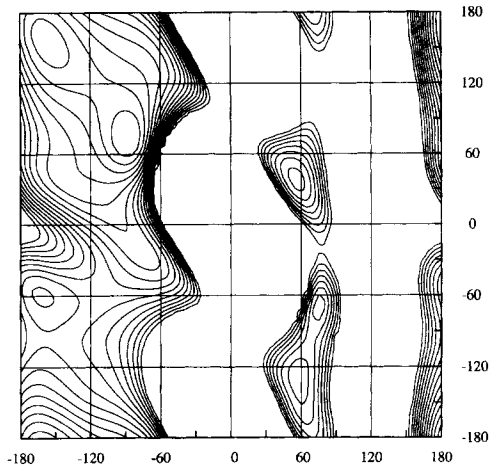

FIGURE 11. Contour plot of the [*rep*] + [*tor*] + [*elec*] energy surface of the model [**merck**] + [**pd**] corrected using a two-dimensional Fourier series of order 6. The contour levels range from 1 to 16 kcal / mol in increments of 1 kcal / mol.

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.[a]

	[**mul**]	[**pd**]	[**md**]	[**mdq**]		
[**charmm**]	0.201	0.192	0.206	0.208	0.233	0.180
[**opls**]	0.059	0.054	0.086	0.056	0.071	0.076
[**amber**]	0.108	0.110	0.107	0.112	0.112	0.100
[**ecepp**]	0.122	0.122	0.113	0.130	0.127	0.092
[**merck**]	0.028	0.034	0.031	0.034	0.029	0.045
[**decomp**]	0.047	0.049	0.058	0.059	0.048	0.064
[**mm3**]	0.199	0.196	0.205	0.208	0.202	0.213

[a] Model 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_{03}, C_{05}) multipole–multipole interaction in the model [[**mdqo**]]. The 49 independent parameters were obtained by minimizing the rmsd over a 24 × 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 ~ .3 kcal/mol. Each such coefficient would create energy differences between conformations of ~ .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 [*elec*] component of the (ϕ, ψ) 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 initio* energy surface for (ϕ, ψ) 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 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 initio* energy surfaces might be accurate enough to enable reliable structure prediction. Any strategy for using a complete set of *ab initio* energy surfaces in the construction of an accurate molecular mechanics model is complicated by the following problems: (1) the possibility of nonnegligible [*pol*], (2) elimination of polarization-induced deformations from atomic charge densities, and (3) accurate representation of [*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 [*elec*], (2) an accurate function of interatomic distances representation of [*rep*], and (3) an accurate two-dimensional Fourier series representation of [*tor*] but neglects (1) [*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 × 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: (ϕ, ψ), (ϕ, χ_1), (ψ, χ_1), and (χ_1, χ_2). The set of conformations over which the (χ_1, χ_2) energy surface is calculated has a single value of (ϕ, ψ), which we will refer to as (ϕ', ψ'). 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 [*rep*], [*tor*], and [*pol*]. This suggests a practical need for two-dimensional corrections to estimates of [*rep*] + [*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 [*rep*] might not be independent of a third adjacent torsion angle degree of freedom.

The (ϕ, ψ) 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 ω, ϕ, and χ_1. Change to ω or χ_1 might greatly reduce a repulsion that depends on distance as the inverse twelfth power. A two-dimensional Fourier series that compensates on the (ϕ, ψ) 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 (ω, ϕ) and (ϕ, χ_1). In general, any contribution to [*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 (ϕ, ψ) energy-surface contour plots.

References

1. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
2. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.*, **7**, 230 (1986).
3. J. C. Smith and M. Karplus, *J. Am. Chem. Soc.*, **114**, 801 (1992).
4. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).
5. J.-H. Lii and N. L. Allinger, *J. Comp. Chem.*, **12**, 186 (1991).
6. T. A. Halgren, *J. Am. Chem. Soc.*, **114**, 7827 (1992).
7. J. R. Maple, M.-J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler, *J. Comp. Chem.*, **15**, 162 (1994).
8. N. L. Allinger, Y. H. Yuh, and J.-H. Lii, *J. Am. Chem. Soc.*, **111**, 8551 (1989).
9. M. J. Dudek and H. A. Scheraga, *J. Comp. Chem.*, **11**, 121 (1990).
10. R. E. Bruccoleri and M. Karplus, *Biopolymers*, **26**, 137 (1987).
11. R. E. Bruccoleri, E. Haber, and J. Novotný, *Nature*, **335**, 564 (1988).
12. R. M. Fine, H. Wang, P. S. Shenkin, D. L. Yarmush, and C. Levinthal, *Proteins*, **1**, 342 (1986).
13. J. Moult and M. N. G. James, *Proteins*, **1**, 146 (1986).
14. U. Berkert and N. L. Allinger, *Molecular Mechanics*, ACS Monograph 177, American Chemical Society, Washington, DC, 1982.
15. S. L. Price, *Mol. Simulation*, **1**, 135 (1988).
16. A. J. Stone and S. L. Price, *J. Phys. Chem.*, **92**, 3325 (1988).
17. C. G. Gray, *Can. J. Phys.*, **54**, 505 (1976).
18. A. J. Stone, *Chem. Phys. Lett.*, **83**, 233 (1981).

19. A. D. Buckingham and P. W. Fowler, *J. Chem. Phys.*, **79**, 6426 (1983).
20. A. D. Buckingham and P. W. Fowler, *Can. J. Chem.*, **63**, 2018 (1985).
21. R. McWeeny, *Methods of Molecular Quantum Mechanics*, Academic Press, London, 1989.
22. SPARTAN version 2.0, Wavefunction, Inc., 18401 Von Karman, #210, Irvine, CA 92715.
23. E. Bendetti, In *Peptides*, M. Goodman and J. Meinhofer, Eds., John Wiley & Sons, New York, 1977, p. 257.
24. S. L. Price, J. S. Andrews C. W. Murray, and R. D. Amos, *J. Am. Chem. Soc.*, **114**, 8268 (1992).
25. T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks III, and J. A. Pople, *J. Am. Chem. Soc.* **113**, 5989 (1991).
26. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).
27. R. S. Mulliken, *J. Chem. Phys.*, **23**, 1833 (1955).
28. C. M. Breneman and K. B. Wiberg, *J. Comp. Chem.*, **11**, 361 (1990).
29. M. Tinkham, *Group Theory and Quantum Mechanics*, McGraw-Hill, New York, 1964.
30. R. A. Sack, *J. Math. Phys.*, **5**, 245 (1964).
31. R. A. Sack, *J. Math. Phys.*, **5**, 252 (1964).
32. R. A. Sack, *J. Math. Phys.*, **5**, 260 (1964).
33. G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.*, **87**, 1883 (1983).
34. J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, **111**, 8566 (1989).
35. J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, **111**, 8576 (1989).
36. R. M. Pitzer, *Acc. Chem. Res.*, **16**, 207 (1983).
37. R. F. W. Bader, J. R. Cheeseman, K. E. Laidig, K. B. Wiberg, and C. Breneman, *J. Am. Chem. Soc.*, **112**, 6530 (1990).
38. M. Waldman and A. T. Hagler, *J. Comp. Chem.*, **14**, 1077 (1993).
39. S. L. Price, C. H. Faerman, and C. W. Murray, *J. Comp. Chem.*, **12**, 1187 (1991).
40. W. A. Sokalski and A. Sawaryn, *J. Chem. Phys.*, **87**, 233 (1987).