Derivation of Class II Force Fields. I. Methodology and Quantum Force Field for the Alkyl Functional Group and Alkane Molecules

J. R. MAPLE, M.-J. HWANG, T. P. STOCKFISCH, U. DINUR,† M. WALDMAN, C. S. EWIG, and A. T. HAGLER*

Biosym Technologies, Inc., 9685 Scranton Road, San Diego, California 92121

Received 29 July 1993; accepted 19 August 1993

ABSTRACT

A new method for deriving force fields for molecular simulations has been developed. It is based on the derivation and parameterization of analytic representations of the *ab initio* potential energy surfaces. The general method is presented here and used to derive a quantum mechanical force field (QMFF) for alkanes. It is based on sampling the energy surfaces of 16 representative alkane species. For hydrocarbons, this force field contains 66 force constants and reference values. These were fit to 128,376 quantum mechanical energies and energy derivatives describing the energy surface. The detailed form of the analytic force field expression and the values of all resulting parameters are given. A series of computations is then performed to test the ability of this force field to reproduce the features of the *ab initio* energy surface in terms of energies as well as the first and second derivatives of the energies with respect to molecular deformations. The fit is shown to be good, with rms energy deviations of less than 7% for all molecules. Also, although only two atom types are employed, the force field accounts for the properties of both highly strained species, such as cyclopropane and methylcyclopropanes, as well as unstrained systems. The information contained in the quantum energy surface indicates that it is significantly anharmonic and that important intramolecular coupling interactions exist between internals. The representation of the nature of these interactions, not present in diagonal, quadratic force fields (Class I force fields), is shown to be important in accounting accurately for molecular energy surfaces. The Class II force field derived from the quantum energy surface is characterized by accounting for these important intramolecular forces. The importance of each

*Author to whom all correspondence should be addressed.
†Present address: Dept. of Chemistry, Ben Gurion Univ. of the Negev, Beersheva 84105, Israel.

Journal of Computational Chemistry, Vol. 15, No. 2, 162–182 (1994)

CCC 0192-8651/94/020162-21

of the interaction terms of the potential energy function has also been assessed. Bond anharmonicity, angle anharmonicity, and bond/angle, bond/torsion, and angle/angle/torsion cross-term interactions result in the most significant overall improvement in distorted structure energies and energy derivatives. The implications of each energy term for the development of advanced force fields is discussed. Finally, it is shown that the techniques introduced here for exploring the quantum energy surface can be used to determine the extent of transferability and range of validity of the force field. The latter is of crucial importance in meeting the objective of deriving a force field for use in molecular mechanics and dynamics calculations of a wide range of molecules often containing functional groups in novel environments.

Introduction

The rapid pace of experimental research in complex molecular systems, particularly in biochemistry and biophysics, has led to greatly increased need for accurate molecular simulation techniques. The most powerful techniques at present are molecular mechanics and dynamics, which presuppose an analytic energy function of the atomic coordinates, the force field, that then governs all the relevant molecular energetic, structural, and dynamic properties. Although in some applications only a qualitative picture derived from the simulation is sufficient, in the great majority of cases an increasingly high level of accuracy is critical for obtaining meaningful results. Examples include refinement of X-ray crystallographic[1] and nuclear magnetic resonance (NMR)[2] structural models of biological species, simulation of enzymatic reaction rates[3] and binding constants,[4] and prediction of protein structures and interaction energies.[5]

A number of factors affect the accuracy of molecular simulations as compared to practical experimental results, including the treatment of solvent, counterions, and other species that may be present in the modeled environment, as well as other details of the simulation techniques. However, no advance in these techniques can compensate for inadequacies in the underlying force field. Further, because in practice the final results will generally depend on a delicate balance of many differing types of terms in the model, inadequacies in a force field may be manifested by gross qualitative errors in predicted results. It is therefore critical that there exist an accurate and systematic procedure for deriving the force fields needed to model any molecular species or environment of interest. Demonstrating and testing such a procedure is the topic of this article and others that will follow in this series.[6]

Procedurally, although force fields have traditionally been derived almost exclusively from experimental data,[7–25] including thermodynamic properties, vibrational frequencies, gas-phase molecular structures, and crystal structures, a lack of such experimental data makes it difficult and in some cases impossible to parameterize and test accurate potential energy functions for use in molecular mechanics and dynamics.[26,27] For example, there is little experimental information on the gas-phase structures, vibrational frequencies, and energetic properties of amino acids, ammonium cations, and carboxylate anions. Thermodynamic data, such as conformational energy differences, rotational barriers, and sublimation energies, are especially useful for developing force fields and are also rare for many functional groups. In addition, experimental structures and spectroscopic information on higher-energy conformers or transition states are relatively nonexistent.

Further, because our major objective in carrying out theoretical molecular simulations of biological and other organic systems is to study molecules for which detailed structural or thermodynamic information is not available from experiment, we desire a predictive tool, and hence force-field functional forms and parameters consequently must be transferable. To derive a transferable force field, it is generally necessary to fit substantial amounts of data with the number of independent experimental observables far exceeding the number of parameters, thereby enabling extrapolation or transferability to molecular environments not included in the set of molecules from which the force constants are determined. Unfortunately, because for most functional groups a sufficient amount of this experimental data is not available, the ratio between the numbers of observables and parameters can be

increased only by reducing the number of parameters. This essentially means using a smaller number of force constants from a simplified and correspondingly less accurate and less transferable potential energy function. Thus, the problem of limited experimental data has been one of the fundamental limitations both in deriving parameters and in developing improved potential energy functional forms.

Recently, a method for exploiting the vast amount of information that can be obtained from *ab initio* quantum mechanical calculations to, in a way, amplify the experimental data was introduced.[26,27] It not only results in a considerable increase in the ratio of observables to parameters, but it also provides an objective criteria for the testing of alternative functional forms and the transferability of force constants. The method makes use of distorted structures of molecules created by deformations of all internal coordinates. The energy and first and second derivatives of the energy with respect to the atomic Cartesian coordinates are calculated for each of these distorted structures by *ab initio* molecular orbital methods. These derivatives provide literally thousands of data that describe the quantum mechanical energy surface of the molecule. For M structures of a molecule containing N atoms, the resulting *ab initio* data include $M - 1$ relative energies, $M(3N - 6)$ independent first derivatives, and $M(0.5)(3N - 6)(3N - 5)$ second derivatives of the energy with respect to a well-defined set of internal coordinates. For example, only five structures of *n*-butane, which has 14 atoms, yield a total of 3514 independent quantum mechanical "observables." This is only a single, relatively small, molecule. Also, when parameterizing a quantum energy surface numerous differing molecules are used, as shown below, thereby providing more than enough information for parameterizing complex potential energy functions. Further, to account for the intricacies of the energy surfaces of molecules, this data contains a wealth of information on the energy of distorted structures (i.e., anharmonicity), transition states, and the energy changes accompanying displacements of particular internal coordinates, information that is virtually inaccessible to experiment.

It is important to note that parameterizing the energy of a molecule employing the quantum mechanical energy surface not only provides a means for determining the numerical values of individual terms, but also permits development of the *analytic form* of the energy expression as well.[26] Differing types of terms in the energy expression may be added, deleted, or reformulated using the fit to the theoretical energy surface and its derivatives as criteria of physical importance.[26] Thus, the complete energy surface provides in a sense a laboratory for developing increasingly more realistic and powerful types of force fields.

This article describes the derivation of a quantum mechanical force field (QMFF) that approximates the Hartree–Fock (HF) 6-31G* potential energy surface of alkanes, including highly strained small ring molecules (e.g., cyclopropane and cyclobutane). The methodology (previously used to derive a quantum mechanical force field for the formate anion[26,27]) is extended to fit the *ab initio* potential energy surfaces of a representative set of alkanes, thereby maximizing the transferability of the resulting force field to all saturated hydrocarbons, regardless of the nature of branching or cyclization or the degree of strain energy. The accuracy of the QMFF potential energy function is assessed with regard to the fit of the quantum mechanical potential energy surface. In particular, the importance of including anharmonicity and cross-terms, as reflected by the energy surfaces of organic molecules, is reexamined for the case of alkanes. The importance of these terms has been a subject of several previous studies.[26–30]

The first section of this article describes a method for sampling the *ab initio* potential energy surface with distorted structures generated by displacing atoms in the directions of the normal modes of vibration. The next section defines the characteristics of the functional form that emerge from the fit to this energy surface. Next, a least-squares method for parameterizing potential energy functions to fit *ab initio* energy and energy derivative data is described, along with the derivation of the force field. Then, the quality of fit to the *ab initio* energy surface is presented. Finally, the importance of anharmonic terms and cross-terms derived from the detailed form of the *ab initio* potential energy surface of alkanes is also examined.

Force fields derived by this method provide a good starting point for applications to large molecules when experimental data for (relatively small) representative molecules of a given functional group is too limited or even entirely unavailable. This method is also useful when one needs a prompt derivation of necessary terms in the force field and time is not available for fits of experimental data. The resulting force fields should provide similar quality to that which would be ob-

tained if quantum mechanical calculations could be used on the complete system of interest. However, because systematic errors are introduced into the force field by the quantum mechanical data (e.g., stretching force constants from HF are known to be ~15% too large),[31,32] a method for refining the potential energy function to fit experimental data is needed and is described in a following article.[6] This article describes only the first part of a two-part method for deriving force fields from, first, quantum mechanical data and, second, experimental data.[6] Of course, this two-step methodology should be most valuable for those functional groups for which only a limited amount of useful experimental data is available (i.e., the great majority of functional groups). Nevertheless, comparisons with alkane force fields previously derived exclusively from experimental data[7-23] assist in the evaluation of the methodology and the establishment of a benchmark for the evaluation of alternative functional forms, as well as of force fields for other functional groups. This two-step

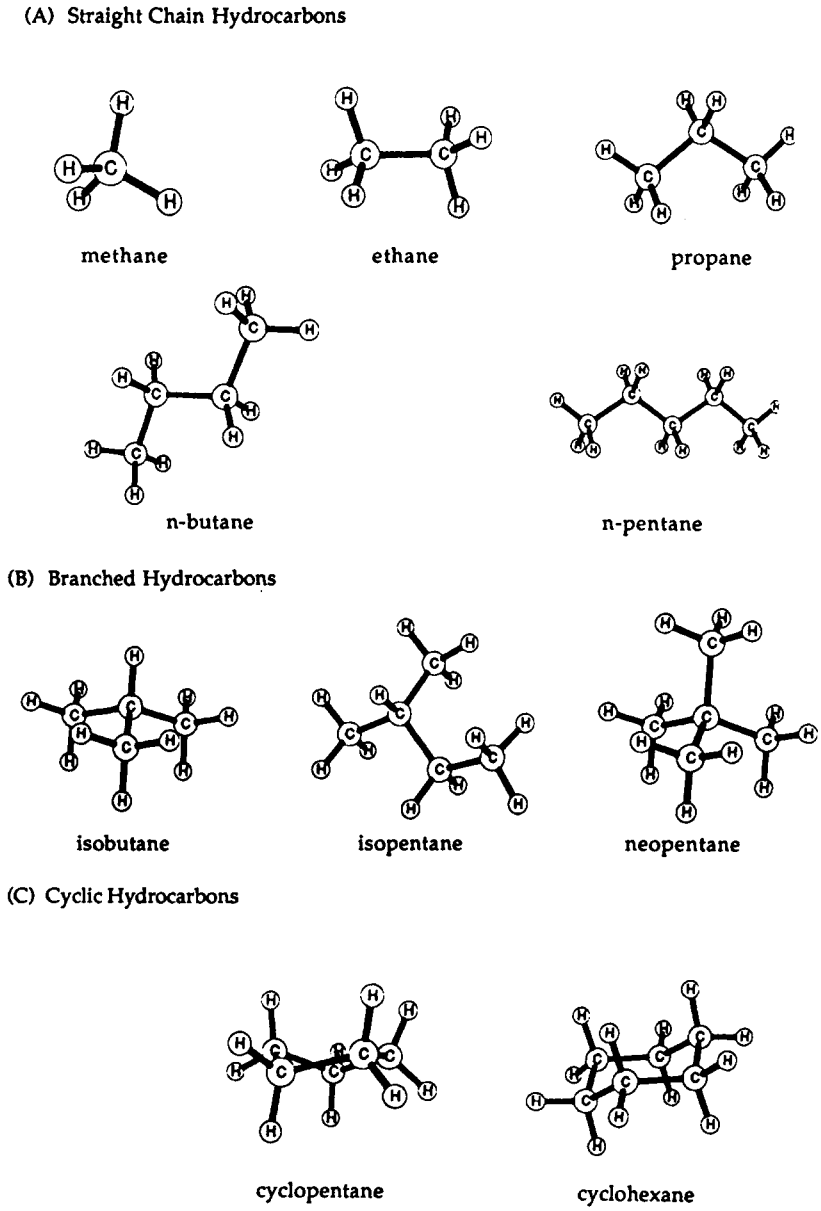


FIGURE 1. (A) Straight-chain, (B) branched, (C) cyclic, (D) four-membered ring, and (E) three-membered ring molecules used to comprise the training set.

methodology for deriving force fields can be considered a logical extension of Lifson's consistent force field (CFF) method[11–15] for deriving force fields from as wide a variety of experimental data as possible because it first takes advantage of the vast amount of data available from quantum mechanics and then uses the same wide variety of experimentally available data, as proposed in the CFF methodology.

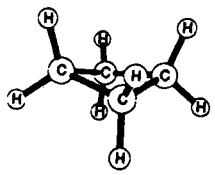
Potential Energy Surface Sampling

Several steps are involved in sampling the potential energy surface. First, a training set of molecules must be selected for the parameterization. Then, the normal modes of vibrational displacements are computed from an equilibrium structure derived by energy minimization. Next, the atoms are displaced in the directions of the normal mode vector components. Finally, the energy, forces, and curvature of the *ab initio* potential energy surface are computed for the distorted structures.

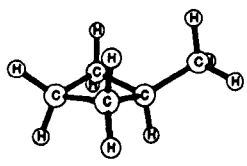
Molecules chosen for the training set are depicted in Figure 1. Methane, ethane, propane, *n*-butane, and *n*-pentane were selected to represent straight-chain *n*-alkanes, while isobutane, isopentane, and neopentane were chosen to give information on branched-chain hydrocarbons. Cycloalkanes were sampled by including cyclopentane and cyclohexane, and cyclobutane, methylcyclobutane, cyclopropane, and methyl cyclopropane derivatives were included to ensure the sampling of highly strained interactions, as discussed below.

Equilibrium structures and normal modes of vibration of the alkanes in the training set were generated by DISCOVER (Biosym Technologies), a molecular mechanics and dynamics program employing the CVFF empirical force field,[13–15] or with the AM1 Hamiltonian in the semiempirical programs MOPAC or AMPAC (QCPE). Gaussian 88[33] (Gaussian, Inc.), an *ab initio* quantum mechan-

(D) Four-membered Rings

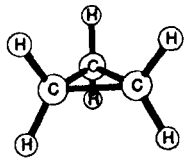


cyclobutane

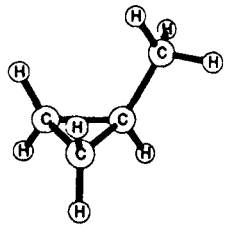


methylcyclobutane

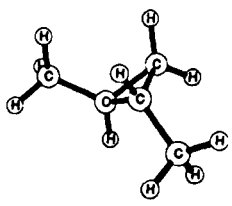
(E) Three-membered Rings



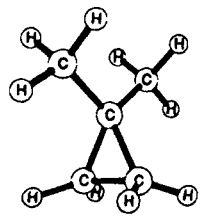
cyclopropane



methycyclopropane



1,2-dimethylcyclopropane



1,1-dimethylcyclopropane

FIGURE 1. (Continued)

ics program, was also used to optimize molecular structures. For distorted structures, GRADSCF (Polyatomics, Inc.) was employed to compute the energies and energy derivatives. Because the goal was to create random distorted structures that explore the quantum mechanical energy surface, it does not matter in principle which of these programs is used to calculate the starting geometry and normal modes. However, *ab initio* calculations are of course computationally more expensive.

To determine the amplitude of atom displacement in the directions of the normal modes of vibration, the energy of *each individual normal mode* was randomly selected within a range of 0–2 kcal/mol, generally the range of interest in molecular mechanics and molecular dynamics. Then, the displacement amplitude (ΔQ_ω) in mass-weighted coordinates was computed with the harmonic approximation[34]

$$E_\omega = 0.5\omega^2(\Delta Q_\omega)^2 \qquad (1)$$

from the vibrational frequency (ω) in rad/s and the randomly selected energy (E_ω) of each normal mode. Next, the representation of the orthogonalized displacement vectors was transformed from Cartesian to internal coordinates, and each component of these normalized displacement vectors was multiplied by the random amplitude (ΔQ_ω) of the mode. Each component of the resulting linearly independent displacement vectors displaced an internal coordinate, and the corresponding components for each mode were simultaneously added to the internal coordinates of the equilibrium structure. Finally, Cartesian coordinates were determined from the internal coordinates of the resulting distorted structure. Usually, 5–10 distorted structures were created per molecule. To improve the sampling of torsion potentials, additional structures of ethane, propane, and *n*-butane were created by rigid rotation about the H—C—C—H, H—C—C—C, and C—C—C—C torsions, respectively, starting from the optimized (global minimum) structure. In addition, *ab initio* optimized structures, which were needed for the purpose of testing the force field, were also included in the potential energy surface sample.

Figure 2 gives a set of examples of internal coordinate distortions in methane as generated by this procedure, as well as the resultant energies. The range of distortion is exemplified by these structures. Thus, the sampling for methane includes C—H bond lengths ranging from 1.031 to 1.169 Å and H—C—H angles ranging from 92.8 to 127.8°. The first structure in this figure corresponds to the *ab initio* optimized geometry. All others have

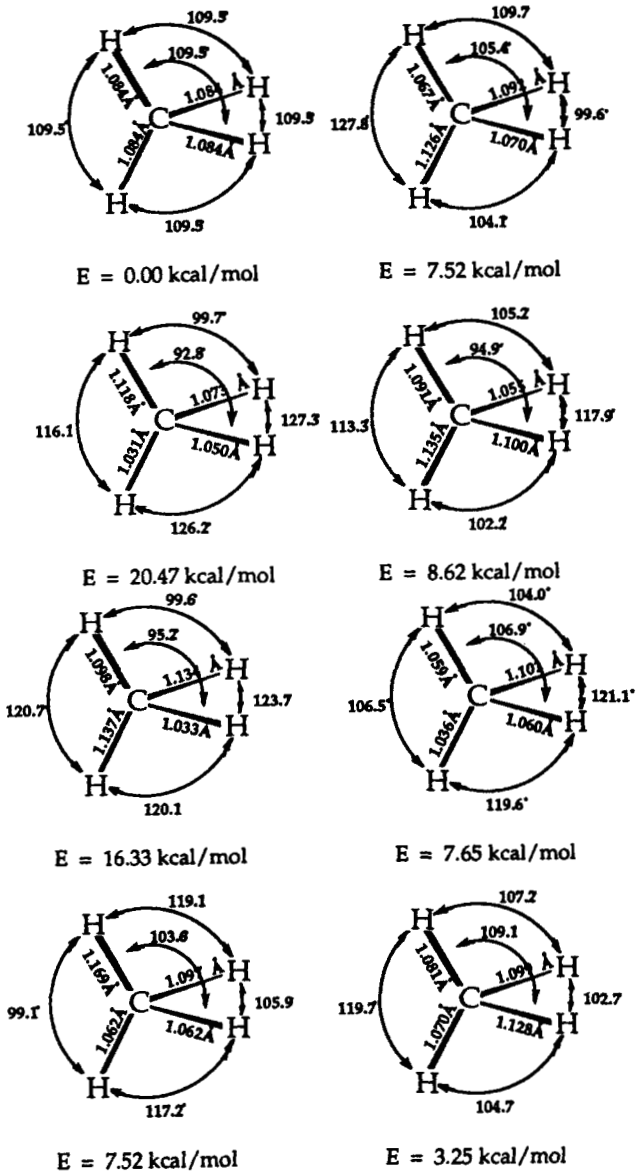


FIGURE 2. Examples of distorted structures used in the derivation of the quantum mechanical energy surface of hydrocarbons. Here, we show several distorted structures of methane from which we calculate relative energies and first and second derivatives. As seen from these structures, the energies range from the minimum energy structure up to 20 kcal/mol. The H—C—H angle is sampled from 93 up to 128° and a large variety of combinations of these angles (to enable sampling of different couplings) are also probed. This is representative of the types of distortions that are generated in the training set of alkane molecules.

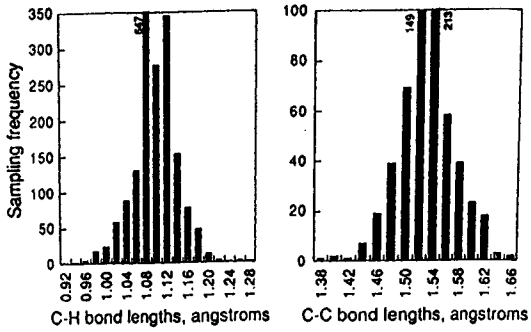


FIGURE 3. Distribution of (left) C—H and (right) C—C bond lengths in structures used to sample the *ab initio* potential energy surface. For the case of C—H bond lengths in the 1.07–1.09 Å, the histogram bar was truncated because the 647 C—H bonds in this range exceeded the dimensions of the ordinate. Two histogram bars for C—C bond lengths were similarly truncated.

been distorted and have *ab initio* energies ranging up to 20.47 kcal/mol relative to the minimum.

The final distorted structures were checked to verify that the internal coordinates were well sampled within a wide range to assure that an adequate portion of the energy surface is explored for predictive purposes. The degree of sampling is depicted in terms of histograms in Figures 3 through 6. From these figures, we see that indeed a large range of distorted geometries are included. Certainly, any distortion we would need to calculate the energetics for a molecular dynamics or mechanics calculation is spanned by these distorted configurations. The C—H and C—C bond lengths are sampled over a range of 0.96 to roughly 1.24 Å for C—H bond lengths and 1.38 to 1.66 Å for C—C

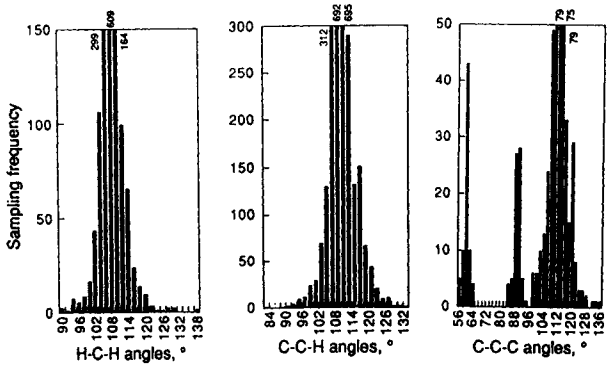


FIGURE 4. Distribution of (left) H—C—H, (middle) C—C—H, and (right) C—C—C bond angles in structures used to sample the *ab initio* potential energy surface.

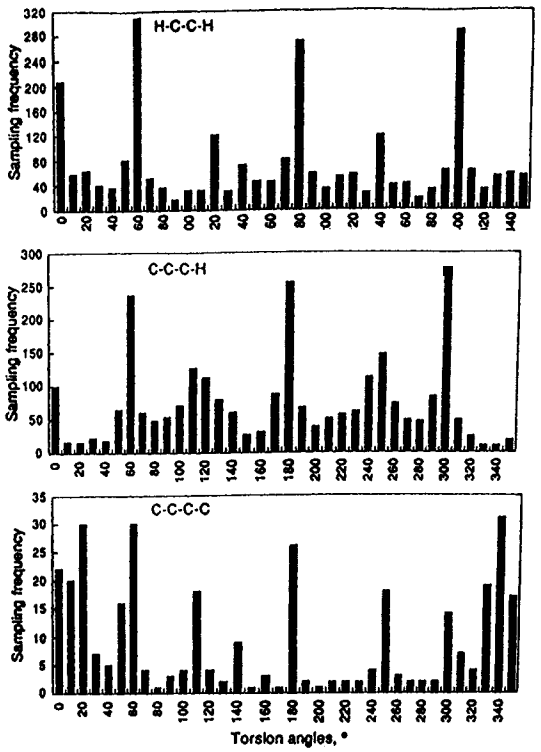


FIGURE 5. Distribution of (top) H—C—C—H, (middle) C—C—C—H, and (bottom) C—C—C—C torsion angles in structures used to sample the *ab initio* potential energy surface.

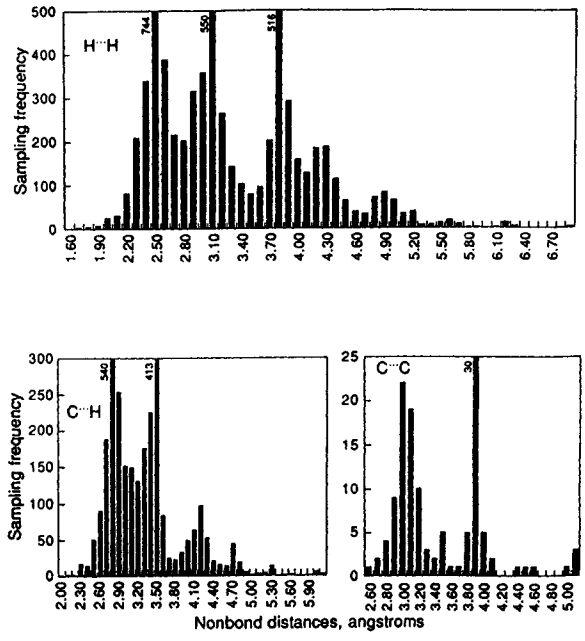


FIGURE 6. Distribution of (top) H···H, (lower left) C···H, and (lower right) C···C distances between nonbonded atoms in structures used to sample the *ab initio* potential energy surface.

bond lengths. Likewise, the H—C—H angle is sampled over a range of almost 50°, and the C—C—H angle from 84 to 132°. One can also see that these are roughly normal curves and most of the internals are sampled at about what would be expected in their equilibrium nonperturbed structures. Figure 3, for example, shows that the final set of distorted structures for 16 molecules contains 647 C—H bond lengths in the range of 1.07–1.09 Å, while approximately 270 C—H bond lengths were in the range 1.09–1.11 Å within this normal distribution. We note that the torsion angles were sampled in a more uniform distribution to make sure the entire 360° range was covered and that the C—C—C angles have peaks corresponding to the cyclopropane and cyclobutane structures. The small, strained cyclic alkanes were added to the sample set to enable derivation of a single force field that would predict the properties of both unstrained and highly strained alkanes with the same set of parameters and functional form. As Figure 5 demonstrates, torsion angles were well sampled throughout the entire 360° range. Especially in the case of the C—C—C—C torsion angles, rigid rotations and energy minimizations at constrained torsion angles were used to improve the sampling of torsion coordinates. Finally, the distances of closest approach between nonbonded atoms (atoms separated by at least three bonds) were limited to 1.6, 2.0, and 2.6 Å for H···H, H···C, and C···C atom pairs, respectively.

Potential energy surface sampling was completed by calculating the *ab initio* energy and the first and second derivatives of the Hartree–Fock 6-31G* energy with respect to the Cartesian coordinates. GRADSCF was used for the energy derivative calculations on distorted structures of methane (8), ethane (42), propane (21), *n*-butane (36), isobutane (17), *n*-pentane (12), isopentane (8), neopentane (6), cyclopropane (8), cyclobutane (10), cyclopentane (14), cyclohexane (9), methylcyclobutane (7), methylcyclopropane (7), 1,2-dimethylcyclopropane (6), and 1,1-dimethylcyclopropane (3). The numbers in parentheses indicate the total numbers of structures, including *ab initio* equilibrium and transition state structures. The resulting total number of independent quantum mechanical observables, including relative energies and first and second derivatives, is 128,376.

The maximum energies, relative to the *ab initio* global minimum, of the distorted structures of the molecules in the training set are listed in the second column of Table I, and the maximum distorted-structure energy divided by the number of degrees of freedom (n_ν) is shown in the third column. Both quantities depend on the extent of distortion of the structure, and the latter quantity is independent of molecular size. The average maximum energy per degree of freedom is about 1.3 kcal/mol for these molecules. The highest value is 2.3 kcal/mol for methane.

QMFF Functional Form

The QMFF functional form was derived by modifying individual terms in the potential energy function and observing the effect on the fit of quantum mechanical data and the accuracy with which *ab initio* equilibrium structures, energies, and energy derivatives were reproduced. The modifications in the analytic representations of the intramolecular force field involved changes in the form of individual energy terms (e.g., harmonic vs. anharmonic bond stretching) or the addition of new types of couplings or cross-terms (e.g., angle/torsion coupling interactions). The effects of these functional form modifications were observed for fits of quantum mechanical data, and the functional form alterations that result in the greatest improvements in the fits of the *ab initio* potential energy surface were retained in the QMFF functional form.

TABLE I.
Maximum Relative Energies (kcal/mol) of Distorted Structures (E_{max}) and the Maximum Relative Energies Divided by the Number of Degrees of Freedom (n_ν) Describing the Hartree–Fock Energy Surface of Hydrocarbons.

Molecule	E_{max}	E_{max}/n_ν
Methane	20.5	2.3
Ethane	31.4	1.7
Propane	55.4	2.1
n-Butane	67.3	1.9
n-Pentane	28.0	0.6
Isobutane	68.0	1.9
Isopentane	20.6	0.5
Neopentane	21.4	0.5
Cyclopentane	70.1	1.8
Cyclohexane	35.9	0.7
Cyclobutane	21.8	0.7
Methylcyclobutane	61.5	1.6
Cyclopropane	21.3	1.0
Methylcyclopropane	41.6	1.4
1,2-Dimethylcyclopropane	52.7	1.1
1,1-Dimethylcyclopropane	48.2	1.0

The individual terms of the QMFF functional form are described below and compared, where appropriate, to previous force fields often employed in molecular mechanics and dynamics, particularly AMBER,[24] CHARMM,[25] and CVFF,[12–15] as well as MM2[19,20,35] and MM3.[21,35] For this comparison, it is useful to classify these various force fields in terms of their functional forms. Thus, AMBER and CHARMM, which are simple diagonal quadratic force fields, are termed "Class I." CVFF and MM2 contain some limited coupling terms and anharmonicity but are still necessarily restricted in their functional forms. MM3 contains extensive anharmonicity and coupling. MM3 and CFF93[6] (based on QMFF), which contain a comprehensive set of anharmonic and coupling terms, will be referred to as "Class II." The detailed definition of Class II force fields is given in Hwang et al.[6]

BOND STRETCHING

In the resulting QMFF force field, the diagonal potential energy function for bonds is the fourth-degree polynomial

$$E_b = {}^2K_b(b - b_0)^2 + {}^3K_b(b - b_0)^3 + {}^4K_b(b - b_0)^4 \quad (2)$$

where b is the bond length and b_0 is the reference value. By contrast, the AMBER and CHARMM force fields employ only the first, the simple harmonic, term, and MM2 only the first two. CVFF uses a Morse potential, which is roughly equivalent but computationally less efficient. It is well known that bond energies are highly anharmonic[14,36] and that anharmonicity is well represented by cubic and quartic terms.[26,27] The quartic term, which is also employed in the MM3 force field, prevents bonds from breaking during molecular mechanics or dynamics calculations, which can occur when only the cubic term is used.

ANGLE BENDING

A fourth-degree polynomial

$$E_\theta = {}^2K_\theta(\theta - \theta_0)^2 + {}^3K_\theta(\theta - \theta_0)^3 + {}^4K_\theta(\theta - \theta_0)^4 \quad (3)$$

is also used to describe the anharmonicity in the bond angle (θ) energy function. Again, AMBER[24] and CHARMM[25] rely on only a simple harmonic approximation, as does CVFF, while MM2 has harmonic and sixth-order terms. This functional form is still substantially less elaborate than the sixth-degree polynomial used in the MM3 formulation. As noted in the following sections, the cubic and quartic term in eq. (3) represent anharmonicity in the angle bending. This anharmonicity was manifested by the energy surface of the highly strained molecules with three- and four-memberd rings.

TORSION

The potential energy function for torsion angles ϕ is given by the three-term Fourier expansion

$$E_\phi = {}^1K_\phi(1 - \cos\phi) + {}^2K_\phi(1 - \cos 2\phi) + {}^2K_\phi(1 - \cos 2\phi) \quad (4)$$

All three cosine terms are needed to fit the conformational energies of hydrocarbons.[20,37] The periodic cosine function is employed by essentially all common types of force fields, although in most of them two of the three terms in the expansion are generally omitted. Although the phases in the first and third terms in eq. (4) differ from the MM3 formulation,[21] both energy functions are equivalent and differ only by a constant term that contributes to absolute, but not relative, energies of a given molecule.

NONBOND

For nonbond interactions between atoms i and j (separated by at least three bonds and by a distance r), the QMFF nonbond potential energy function is an inverse-power expression:

$$E_n = q_iq_j/r + \varepsilon[2(r^*/r)^9 - 3(r^*/r)^6] \quad (5)$$

An inverse 9 power was found to be necessary to fit both intramolecular and intermolecular properties simultaneously including rotational barriers and fits to crystal structures. A similar suggestion was hypothesized by Lifson some years ago, suggesting that the 12th-power repulsion was too "hard." More recently, Allinger has also concluded that the simple 12th-power repulsion is counterindicated by the combination of intra- and intermolecular properties and has used a softer exponential term in MM3. Class I force fields including CVFF, AMBER, and CHARMM, of course, all use inverse 12th-power repulsions. We also employ a recently introduced set of combination rules for relating the individual atomic parameters to the pairwise potential[38]:

$$r^* = [(r_i^{*6} + r_j^{*6})/2]^{1/6} \quad (6)$$

and

$$\varepsilon = (\varepsilon_i\varepsilon_j)^{1/2}2(r_i^*r_j^*)^3/[r_i^{*6} + r_j^{*6}] \quad (7)$$

Here, q_i, ε_i, and r_i^* are the partial charge, well depth, and diameter of atom i. Thus, the potential function in eq. (5) uses Coulomb's law to approximate the electrostatic interaction, an inverse 9th-power repulsive interaction between atoms, and an attractive dispersion interaction. The new combination rules in eqs. (6) and (7) have been shown to reproduce experimental rare gas data to a much better extent than more commonly used combination rules (e.g., arithmetic and geometric means) used in previous force fields.[38]

BOND INCREMENTS

The partial atomic charges in eq. (5) are computed from the bond incremental charges or bond increments. The bond increment δ_{ij} is defined as the partial charge contributed by atom j to atom i.[39] In general, each of the n atoms bonded to an atom contributes to the atomic charge. By computing a partial atomic charge for atom i from the summation over the bond increments

$$q_i = \sum_{j=1}^{n} \delta_{ij} \tag{8}$$

which depend on the relative electronegativity of the bonded atoms, the dependence of the partial charge on the bonding environment is automatically taken into account. Consequently, bond increments can be more transferable than partial atomic charges.[39] An additional advantage of using bond increments is that charge conservation is automatic. When charges are directly transferred to another molecule, the sum of partial atomic charges in that molecule is, in general, not equal to the charge of the molecule (i.e., zero). However, this problem does not occur when bond increments are used to compute the atomic charges.

The first atom in the bond increment subscript is defined as the charge acceptor and the second as the charge donor. This convention is expressed by the relationship

$$\delta_{ij} = -\delta_{ji} \tag{9}$$

between all bonded atom pairs i and j. Equation (9) states that the charge donated by atom j to atom i is opposite in sign to that donated by i to j. Consequently, only a single bond increment parameter is needed to describe each unique pair of bonded atoms. For alkanes, this means that the δ_{CH} bond increment is the only adjustable electrostatic parameter. The δ_{CC} bond increment is fixed at zero because two carbon atoms bonded to each other have the same tendency to donate or accept electrons.

BOND/BOND COUPLING

The QMFF functional form for bond/bond cross-terms is

$$E_{bb'} = K_{bb'}(b - b_0)(b' - b_0') \tag{10}$$

and is identical in form to the cross-term used by CVFF.[14,15] The inclusion of this bond/bond coupling interaction, which is neglected by MM3, is necessary for a force field to reproduce vibrational frequencies with high accuracy.[15,26,27,37] Bond/bond coupling was revealed from fits of *ab initio* second derivatives of the energy with respect to Cartesian coordinates, as well as from calculations of the second derivatives of the energy with respect to bond lengths. Although this coupling interaction is significant in alkanes, the second derivatives in other species reveal that the coupling is particularly strong for bonds in systems with delocalized electrons (e.g., the CO/CN bond coupling in amides).

Following the convention used by CVFF,[14,15] only bonds with an atom in common are coupled, as depicted in Figure 7, which illustrates bond/bond interactions (Fig. 7a), as well as the other cross-term interactions used in the QMFF force field. The neglect of coupling of bonds without an atom in common is ordinarily a good approximation.[40] During the derivation of QMFF for other species from fits of *ab initio* potential energy surfaces, the only significant exception found for this approximation was for bonds in aromatic ring molecules.

BOND/ANGLE COUPLING

As depicted in Figure 7b, angles are coupled to the two bonds that comprise the bond angle by the energy function

$$E_{b\theta} = K_{b\theta}(b - b_0)(\theta - \theta_0) \tag{11}$$

There are two of these terms for each angle (i.e., one for each bond). Bond/angle coupling is necessary to reproduce frequencies[14,26,27,37] and several structural features. For example, in tri-tert-butylmethane,[14] as well as in other molecules, bond/angle coupling is needed to reproduce the bond lengthening that is observed when the bond angle is compressed in strained molecules, as well as to account for the concomitant shifts in vibrational frequencies.

The neglect of coupling of an angle to other bonds (i.e., bonds other than the two that define the bond angle) is ordinarily a good approximation,[40] and this has been verified by fits of *ab initio*

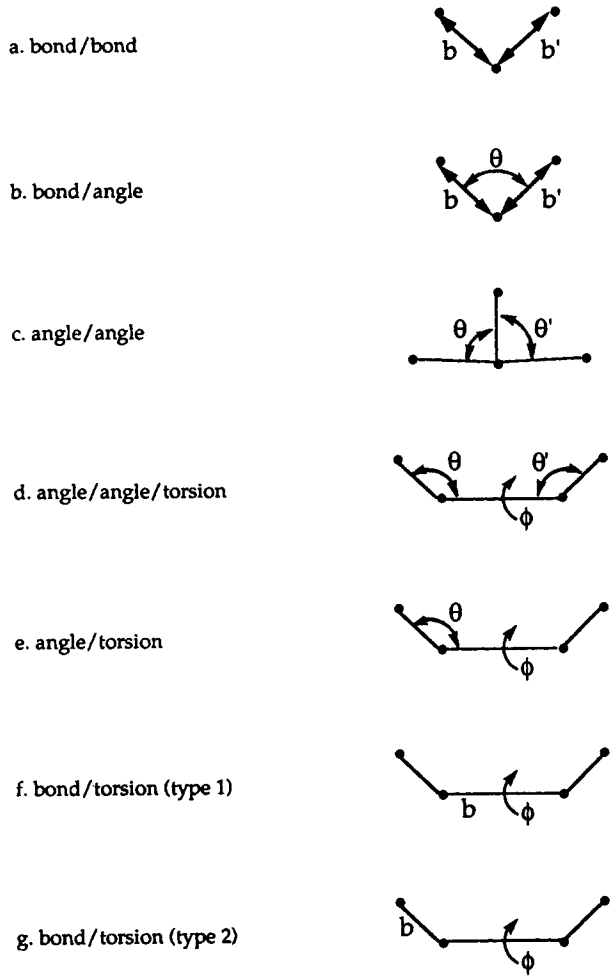


FIGURE 7. Representation of cross-terms (coupling interactions) in QMFF.

potential energy surfaces of a variety of functional groups. This approximation, as well as the functional form in eq. (11), were also used in CVFF[14,15] as well as MM2 and MM3.[21,37] None of the cross-terms were present in the CHARMM and AMBER force fields, which are diagonal quadratic functional forms. Thus, consideration of these force fields is omitted in discussing all subsequent coupling interactions.

ANGLE/ANGLE COUPLING

Bond angles are coupled to other bond angles that share a common bond, as illustrated in Figure 7c, and the QMFF functional form is given by

$$E_{\theta\theta'} = K_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0') \qquad (12)$$

This interaction has been used in the CVFF[14,15] and MM3[21] force fields to accurately reproduce experimental vibrational frequencies for coupled bending modes. However, it has been suggested that this term is relatively unimportant for bond angles in tetracoordinate centers.[40] The results obtained from the fit of the quantum energy surface, especially for the relative energies and forces in molecules with large amounts of angle strain (i.e., cyclopropane and methyl derivatives of cyclopropane), support the importance of the angle/angle coupling interactions.

ANGLE/ANGLE/TORSION COUPLING

The QMFF functional form for the angle/angle/torsion cross-term is

$$E_{\theta\theta'\phi} = K_{\theta\theta'\phi}(\theta - \theta_0)(\theta' - \theta_0')\cos\phi \qquad (13)$$

As indicated in Figure 7d, this term describes the coupling between a torsion angle *A-B-C-D* and the two vicinal bond angles *A-B-C* and *B-C-D* that share the common bond *B*—*C*. This coupling term, which was first introduced by Lifson and Warshel,[41] is used by CVFF to more accurately reproduce vibrational frequencies,[13–15] but it is not used by MM3. It is found that this coupling interaction makes a large contribution to the *ab initio* relative energies, forces, and coupling derivatives in all alkanes (except methane) examined here (see Table I). It is most notable in cyclobutanes, wherein it accounts for a significant proportion of the ring pucker and inversion barrier. Similar results were noted from fits of *ab initio* data of all other functional groups that have been examined. Thus, information implicit in the quantum energy surface supports Lifson and Warshel's conclusion as to the importance of this term.

BOND/TORSION AND ANGLE/TORSION COUPLING

The bond/torsion and angle/torsion interactions shown in Figures 7e–7g are expressed as the displacement of a bond length or bond angle multiplied by a Fourier expansion in the torsion coordinate, as indicated in eqs. (14) and (15).

$$E_{\phi b} = (b - b_0)[{}^1K_{\phi b}\cos\phi + {}^2K_{\phi b}\cos 2\phi + {}^3K_{\phi b}\cos 3\phi] \qquad (14)$$

$$E_{\phi\theta} = (\theta - \theta_0)[{}^1K_{\phi\theta}\cos\phi + {}^2K_{\phi\theta}\cos 2\phi + {}^3K_{\phi\theta}\cos 3\phi] \qquad (15)$$

As shown in Figure 7e, there are two angle/torsion interactions for every torsion angle ϕ. There are also two types of bond/torsion couplings. The first

type (in Fig. 7f) involves the coupling of the central bond to the torsion angle, while the second type (in Fig. 7g) couples the peripheral bond to the torsion angle. There are two bond/torsion interactions of the second type for every torsion angle because there are two peripheral bonds. The CVFF force field does not employ angle/torsion coupling or bond/torsion coupling, and MM3 uses only the first of these bond/torsion interactions.

Differentiation of eq. (14) with respect to the bond length gives

$$\partial E_{\phi b'}/\partial b = {}^1K_{\phi b} \cos \phi + {}^2K_{\phi b} \cos 2\phi + {}^3K_{\phi b} \cos 3\phi \quad (16)$$

which shows that this interaction describes the dependence of the bond length restoring force on the torsion coordinate ϕ. Similarly, the angle/torsion coupling term is used to determine the torsion angle dependence of the bond angle restoring force, that is, these interactions are important for describing the structure of molecules in which different conformers exhibit significant differences between bond lengths or bond angles. The bond/torsion interaction is reflected in the bond stretching that occurs with eclipsed bonds (i.e., *cis* torsion angles) in structures such as eclipsed ethane. The fits of the alkane *ab initio* energy surface (particularly the energies and forces) reveals the existence of the bond/torsion and angle/torsion coupling interactions, both for alkanes as well as acetals, carbohydrates, amides, and other functional groups. The bond/torsion and angle/torsion coupling terms are found to reproduce the anomeric effect in acetals and carbohydrates or other functional groups in which a carbon atom separates two electronegative atoms.[42] It is intriguing that the same bond/torsion and angle/torsion coupling interactions revealed in the hydrocarbon energy surface may explain (or cause) the anomeric effect in carbohydrates. The implications of this are the subject of further study.

The bond/torsion functional form used by MM3 is given by

$$E = k(b - b_0)[1 + \cos 3\phi] \quad (17)$$

and differs from eq. (14). One difference is that the MM3 form lacks the first two cosine terms in the Fourier expansion. Another important difference is the linear term in the MM3 form in eq. (17). This term, which is independent of the torsion angle, is more clearly seen by rewriting eq. (17) as

$$E = k(b - b_0) + k(b - b_0)\cos 3\phi \quad (18)$$

The restoring force on the bond is therefore given by

$$-\partial E/\partial b = -k - k \cos 3\phi \quad (19)$$

Thus, the first term in eq. (18) is a linear term that is effectively added to the polynomial in eq. (2). This linear term introduces a force [i.e., the first term in eq. 19)] on the bond length that is *independent* of the torsion angle, thereby driving the bond length away from the idealized (strain free) bond length (i.e., b_0). The net result of the use of eq. (17) is that k and b_0 are correlated, and the transferability of these two parameters is reduced. This problem is eliminated by the use of the functional form in eqs. (14) and (15).

The full functional form that describes the quantum energy surface is given by the sum of the individual terms in eqs. (2)–(15). The result is

$$\begin{aligned}
E = {} & \sum_b [{}^2K_b(b - b_0)^2 + {}^3K_b(b - b_0)^3 \\
& + {}^4K_b(b - b_0)^4] + \sum_\theta [{}^2K_\theta(\theta - \theta_0)^2 \\
& + {}^3K_\theta(\theta - \theta_0)^3 + {}^4K_\theta(\theta - \theta_0)^4] \\
& + \sum_\phi [{}^1K_\phi(1 - \cos \phi) + {}^2K_\phi(1 \\
& - \cos 2\phi) + {}^3K_\phi(1 - \cos 3\phi)] \\
& + \sum_{i>j} \frac{q_i q_j}{r_{ij}} + \sum_{i>j} \varepsilon\left[2\left(\frac{r^*}{r_{ij}}\right)^9 - 3\left(\frac{r^*}{r_{ij}}\right)^6\right] \\
& + \sum_b \sum_{b'} K_{bb'}(b - b_0)(b' - b_0') \\
& + \sum_\theta \sum_{\theta'} K_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0') \\
& + \sum_b \sum_\theta K_{b\theta}(b - b_0)(\theta - \theta_0) \\
& + \sum_\phi \sum_b (b - b_0)[{}^1K_{\phi b} \cos \phi \\
& + {}^2K_{\phi b} \cos 2\phi + {}^3K_{\phi b} \cos 3\phi] \\
& + \sum_\phi \sum_{b'} (b' - b_0')[{}^1K_{\phi b'} \cos \phi \\
& + {}^2K_{\phi b'} \cos 2\phi + {}^3K_{\phi b'} \cos 3\phi] \\
& + \sum_\phi \sum_\theta (\theta - \theta_0)[{}^1K_{\phi\theta} \cos \phi \\
& + {}^2K_{\phi\theta} \cos 2\phi + {}^3K_{\phi\theta} \cos 3\phi] \\
& + \sum_\phi \sum_\theta \sum_{\theta'} K_{\phi\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0')\cos \phi
\end{aligned} \quad (20)$$

TRANSFERABILITY OF PARAMETERS

The QMFF force field uses the conventional approach of assigning parameters based on atom types. The parameters for two internal coordinates

or cross-terms are distinguished if the atom types that define the internal coordinate or coupling interaction differ. For alkanes, all hydrogen and carbon atoms have atom types of "h" and "c," respectively, irrespective of whether they belong to methyl, methylene, or methine groups. Parameters for methyl, methylene, and methine groups are the same. This is also the case for the CVFF,[14,15] AMBER,[24] and CHARMM[25] force fields. In contrast, the MM3[21] force field uses different bond angle reference values for these groups. Further, the same atom types were used in three-, four-, and five-membered rings, unlike the MM2 and MM3 force fields,[20–23] which employ four different sets of atom types for alkanes (i.e., one each for three-, four-, and five-membered rings plus one more for all other alkanes in addition to the above-mentioned distinction between methyl, methylene, and methine groups). QMFF has been derived with the goal of obtaining the most accurate and physically reasonable energy function and, hence, of maximizing the force-field transferability. This was found to result in a minimal number of atom types (i.e., one set for all strained and unstrained alkanes).

Least-Squares Fit of the Energy Surface

The force constants in eq. (20) were determined with PROBE (Biosym Technologies), a program that uses a least-squares method (with a Levenberg–Marquardt minimizer[43]) to adjust the potential energy function parameters to fit the energy and the first and second derivatives of the energy (with respect to Cartesian atomic coordinates) of one or more molecules and structures. The weighted sum of squared deviations (i.e., the object function to be minimized) was defined as

$$S = \sum_{a,b} W_{ab} \left[W_{eab}[E^r_{ab} - \overline{E}^r_{ab}]^2 + W_{1ab} \sum_{i=1}^{3N_b} \left[\frac{\partial E_{ab}}{\partial x_i} - \frac{\partial \overline{E}_{ab}}{\partial x_i} \right]^2 + \sum_{i \geq j}^{3N_b} W_{2ijab} \left[\frac{\partial^2 E_{ab}}{\partial x_i \partial x_j} - \frac{\partial^2 \overline{E}_{ab}}{\partial x_i \partial x_j} \right]^2 \right] \quad (21)$$

where E_{ab} is the potential energy calculated from eq. (20), and $\overline{E}_{ab}$ is the corresponding quantum mechanical energy of the distorted configuration, a, of molecule b. E^r_{ab} and $\overline{E}^r_{ab}$ are the relative energies of configurations determined from the potential energy function and from quantum mechanics (i.e., the "observables"), respectively. The relative energies for each distorted structure of a molecule, b, are determined relative to the structure with the lowest energy. For the summation over energy derivatives in eq. (21), N_b is the number of atoms in molecule b, and x_i is one of the three N_b Cartesian atomic coordinates.

WEIGHTING FACTORS

The overall weighting factor W_{ab} in eq. (21) allows all data for configuration a of molecule b to be weighted independently. One might, for example, want to weight minima and barriers more heavily than arbitrarily distorted configurations in some cases. For the work presented here, this weighting factor was set to 1.

The weighting factor for the energies W_{eab} was set equal to a value of 15,000. Larger values gave significantly improved fits of the distorted structure energies. However, the predictions of the energy of distorted structures not included in the fit often degraded when larger energy weighting factors were employed. Thus, a value of 15,000 was about the maximum energy weighting factor that could be employed without overfitting the energies of the distorted structures and thereby reducing the transferability of the parameters.

The weighting factor for the first derivatives, W_{1ab}, in eq. (21), was set to 100. Parameterization with values of W_{1ab} ranging from 50 to 200 gave comparable predictions of the *ab initio* equilibrium structures and vibrational frequencies. W_{1ab} values less than 50 resulted in slight improvements in the fit of the curvature of the quantum energy surface ($\partial^2 E/\partial x_i \partial x_j$) and vibrational frequencies, but were accompanied by a significant degradation in first derivative and structure predictions. For values of the weighting factor (W_{1ab}) greater than 200, minor improvements in the fits of first derivatives were accompanied by large errors in the fit of second derivatives of the energy and, consequently, vibrational frequency predictions.

The weighting factor for the second derivatives of the energy is W_{2ijab}. W_{2ijab} was set to 1 for all second derivatives of the energy, with the exception of 1–4 second derivatives, for which a value of 75 was used. This large value was used to compensate for the small size of 1–4 second derivatives relative to 1–2 and 1–3 second derivatives, which are defined as second derivatives with respect to the Cartesian coordinates of atom pairs separated by one or two bonds, respectively. In experiments

experiments in which *ab initio* data for ethane, propane, and *n*-butane were fit separately, substantially different results for the torsion parameters in eq. (5) were determined when W_{2ijab} was less than 75, but they remained constant when W_{2ijab} was increased above a value of 75. For small-ring molecules, some 1–4 second derivatives are simultaneously 1–3 (e.g., some carbon/hydrogen atom pairs in cyclopropane) or 1–2 (e.g., bonded carbon atom pairs in cyclobutane are the 1–4 terminal atoms of a C-C-C-C torsion). These second derivatives were defined by PROBE as 1–2 or 1–3 derivatives rather than 1–4, and were given a weighting factor of 1 because their values were already large.

ORIGIN OF NONBOND PARAMETERS

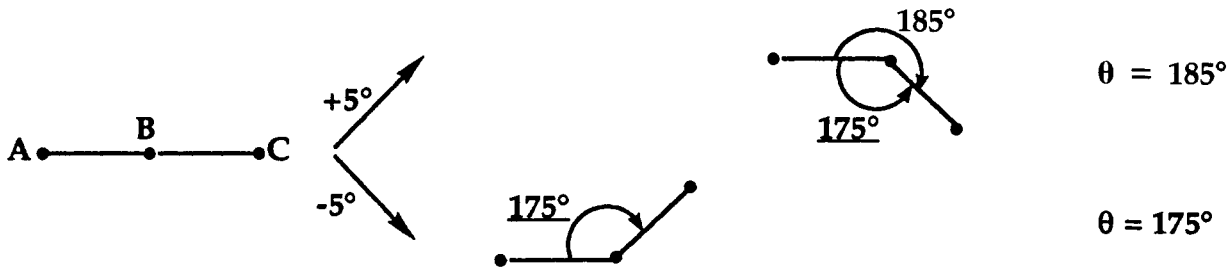
Although charges (or bond increments) and repulsive constants could in principle be determined from fits of the *ab initio* data, a pragmatic approach that served to enhance transferability was employed in which these parameters were obtained from experimental hydrocarbon crystal structures and sublimation energies.[6] These parameters were transferred (see Table II) and held fixed in all least-square fits of the *ab initio* data. The derivation of these parameters will be the subject of a later study.

CONSTRAINTS ON BOND ANGLE ANHARMONIC FORCE CONSTANTS

The cubic and quartic force constants for the bond angle energy function were constrained so that the function in eq. (3) has a slope of 0 at angles of 0 and 180°. This is, in fact, required by symmetry. This can be seen by considering either a 0 or 180° angle. Deformations in either direction by say ±5° produce the same angle between the bonds, i.e., 175 or 185° are equivalent (both 175°). Thus, they have the same energy and the energy function must be symmetric about these 0 and 180° values to reflect this (see Scheme 1). As a result, there were only two independently adjustable parameters (i.e., a reference value and a quadratic force constant) for each of the H-C-H, H-C-C, and C-C-C diagonal energy functions. This procedure also ensures that the polynomial in eq. (3) has only a single minimum in the 0–180° range, as required to prevent unacceptably large or small bond angles during a structure optimization or molecular dynamics simulation. Quadratic force fields (e.g., CVFF, CHARMM, and AMBER) do not have this correct symmetry-required limiting behavior in the bond angle energy function, nor is this limiting behavior obeyed by the MM3 bending energy function.

CONSTRAINTS ON TORSION PARAMETERS

In addition, we tested whether the ${}^1K_\phi$ term for the torsion potential in eq. (4) could be constrained to be the same for the H-C-C-H, H-C-C-C, and C-C-C-C torsions. Constraints on the ${}^2K_\phi$ and ${}^3K_\phi$ force constants were similarly tested. Thus, only three adjustable torsion parameters were used in the fit of the quantum mechanical data. The rationale for testing this constraint is to ask whether, to a first approximation, the potential for twisting about a bond depends primarily on the identity of the two atoms in the bond rather than on the peripheral 1–4 atoms. This would be desirable as it would maximize the transferability of the torsion parameters so that they could be applied to the large number of possible *X*-C-C-*Y* torsions, where *X* and *Y* are any atoms. Only a small degradation in the quality of fit accompanied this reduction in the number of torsion parameters from 9 to 3, and thus we adopted this approximation.



SCHEME 1. Symmetry of angle function about 180°.

Parameterization

A two-step procedure for parameterizing the *ab initio* data was employed. In the first step, data from molecules with a three- or four-membered ring was excluded from the fitting procedure. In the second step, the data from the three- and four-membered rings was added to the parameterization data set, and the parameters in the diagonal bond, bond angle, and torsion angle energy functions were frozen at the values determined from the first step, while all other parameters were redetermined. (This procedure reduced systematic errors in predicted vibrational frequencies.)

Two statistical tests of the parameters were applied. Parameter uncertainties were determined from standard deviations, which were calculated from the diagonal elements of the variance–covariance matrix.[43,44] In addition, parameters were subjected to a *t*-test to check the hypothesis that they were statistically identical to zero.[44] If the probability that this hypothesis was true exceeded 0.05 (at the 95% confidence level), then the hypothesis was accepted and the parameter was fixed at a value of zero during a subsequent parameter refinement. It turned out that this hypothesis was rejected for all adjustable parameters, and so no force constants were set equal to zero in the QMFF alkane force field.

To summarize the fit of the alkane quantum mechanical data, after excluding the 4 nonbond parameters (determined separately from fits of experimental crystal structures and sublimation energies), the 78 remaining force constants and reference values listed in Table II were determined. Only 66 of these parameters were independently adjustable, as described above, because the values of 6 torsion parameters and 6 anharmonic bond angle parameters were determined from symmetry relationships or assumed transferability. Thus, because a total of 128,376 independent *ab initio* observables (i.e., relative energies and the first and second derivatives of the energies) were used an observable-to-parameter ratio of 1945:1 was achieved for the fit of the data.

Quality of Fit to *Ab Initio* Potential Energy Surface

RELATIVE ENERGIES AND ENERGY DERIVATIVES

The quality of fit of the QMFF to the *ab initio* potential energy surface is indicated in Table III, which lists the relative deviations of energies and energy derivatives between those calculated by the QMFF and by *ab initio*. For comparison, Table III also gives deviations between *ab initio* values and those obtained with a harmonic diagonal (Class I) force field, which was derived in a manner similar to that of QMFF (i.e., by eliminating all anharmonic and cross-terms and refitting the harmonic diagonal force constants to the same *ab initio* data). As noted in the discussion of the interaction terms, the quadratic or harmonic diagonal force field is representative of the standard force fields used in the simulation of large organic and biomolecular systems such as AMBER, CHARMM, etc. CVFF, as noted above, contains some coupling terms and anharmonicity, and lies midway between Class I and Class II functional forms in this respect. The second and third columns of Table III give the rms deviations for the relative energies of the harmonic and QMFF force fields. For the Class I function, the deviations range from 0.37 kcal/mol for *n*-pentane to 6.33 kcal/mol for 1,1-dimethylcyclopropane, while for QMFF the deviations range from 0.26 kcal/mol for the distorted structures of methane to 2.87 kcal/mol for methylcyclopropane. The average rms energy deviation, which is given in the last row in Table III, is 0.87 and 2.69 kcal/mol for QMFF and the harmonic diagonal force fields, respectively. This demonstrates that on average the inclusion of anharmonicity and cross-term interactions reduces the errors in calculated energies by 1.8 kcal/mol. The fourth and fifth columns list the same rms deviations as percentages of the maximum relative energy of the distorted structures (listed in Table I). The percentage error in the harmonic force field energies range from 1.3% for *n*-pentane to 17.0% for cyclobutane, while the percentage error for QMFF energies ranges from 0.8% for *n*-butane to 6.9% for methylcyclopropane. The average percentage error over all structures is 2.2 and 6.9% for QMFF and the harmonic diagonal force field, respectively.

The percent rms deviations between the forcefield and *ab initio* first and second energy derivatives were calculated from the ratio between the sum of squared deviations and the sum of squared *ab initio* derivatives. The first derivative errors range from 6.2% for *n*-butane to 19.5% for methylcyclopropane for QMFF while the corresponding range is 15.2% for methane to 65.7% for cyclopropane for the harmonic force field. Similarly, for QMFF the second derivative errors range from 2.6% for ethane to 7.4% for methylcyclopropane, while for the harmonic force field these deviations range from 12.8% for ethane to 24.9% for methylcyclobutane. On the

TABLE II.
Force Constants and Reference Values in QMFF for Alkanes.

A. Bond Length: $E = {}^2K_b(b - b_0)^2 + {}^3K_b(b - b_0)^3 + {}^4K_b(b - b_0)^4$

Bond	b_0 (Å)	${}^2K_b$ (kcal mol^{-1} Å^{-2})	${}^3K_b$ (kcal mol^{-1} Å^{-3})	${}^4K_b$ (kcal mol^{-1} Å^{-4})
H—C	1.0845	417.4	−851.3	1040.0
C—C	1.5297	340.2	−586.1	758.0

B. Bond Angle: $E = {}^2K_\theta(\theta - \theta_0)^2 + {}^3K_\theta(\theta - \theta_0)^3 + {}^4K_\theta(\theta - \theta_0)^4$

Angle	θ_0 (°)	${}^2K_\theta$ (kcal mol^{-1} rad^{-2})	${}^3K_\theta$ (kcal mol^{-1} rad^{-3})	${}^4K_\theta$ (kcal mol^{-1} rad^{-4})
H—C—H	107.7	51.5	−8.9	−10.8
H—C—C	110.8	52.7	−10.9	−11.3
C—C—C	112.9	52.2	−12.1	−11.3

C. Torsion Angle: $E = {}^1K_\phi(1 - \cos\phi) + {}^2K_\phi(1 - \cos 2\phi) + {}^3K_\phi(1 - \cos 3\phi)$

Torsion	${}^1K_\phi$ (kcal mol^{-1})	${}^2K_\phi$ (kcal mol^{-1})	${}^3K_\phi$ (kcal mol^{-1})
H—C—C—H	−1.152	0.012	−0.179
H—C—C—C	−1.152	0.012	−0.179
C—C—C—C	−1.152	0.012	−0.179

D. van der Waals Interaction: $E = \varepsilon[2(r^*/r)^9 - 3(r^*/r)^6]$ where $r^* = [(r_i^{*6} + r_j^{*6})/2]^{1/6}$ and $\varepsilon = (\varepsilon_i\varepsilon_j)^{1/2}2(r_i^*r_j^*)^3/(r_i^{*6} + r_j^{*6})$

Atom	r_i^* (Å)	ε_i (kcal mol^{-1})
H	2.995	0.020
C	4.010	0.054

E. Bond Increment: $E = 332.0\, q_iq_j/r_{ij}$ where $q_i = \Sigma_k\delta_{ik}$

Bond	δ_{ik} (electrons)
H—C	0.053
C—C	0.000

F. Bond/Bond: $E = K_{bb'}(b - b_0)(b' - b_0')$

Bond/Bond	$K_{bb'}$ (kcal mol^{-1} Å^{-2})
H—C/H—C	11.769
H—C/C—C	11.310
C—C/C—C	10.062

G. Bond/Angle: $E = K_{b\theta}(b - b_0)(\theta - \theta_0)$

Bond/Angle	$K_{b\theta}$ (kcal mol^{-1} Å^{-1} rad^{-1})
H—C/H—C—H	22.7
H—C/H—C—C	13.9
C—C/H—C—C	34.5
C—C/C—C—C	18.3

TABLE II.

H. Angle/Angle: $E = K_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0')$

Angle/Angle	Common Bond	$K_{\theta\theta'}$ (kcal mol^{-1} rad^{-2})
H—C—H/H—C—H	H—C	0.93
H—C—H/H—C—C	H—C	1.57
H—C—C/H—C—C	H—C	3.12
H—C—C/H—C—C	C—C	−2.18
H—C—C/C—C—C	C—C	−4.62
C—C—C/C—C—C	C—C	−8.97

I. Angle/Angle/Torsion: $E = K_{\phi\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0')\cos\phi$

Angle/Angle/Torsion	$K_{\theta\theta'\phi}$ (kcal mol^{-1} rad^{-2})
H—C—C/H—C—C/H—C—C—H	−14.8
H—C—C/C—C—C/H—C—C—C	−19.2
C—C—C/C—C—C/C—C—C—C	−35.6

J. Bond/Torsion (type 1): $E = (b - b_0)[{}^1K_{\phi b}\cos\phi + {}^2K_{\phi b}\cos 2\phi + {}^3K_{\phi b}\cos 3\phi]$

Bond/Torsion	${}^1K_{\phi b}$ (kcal mol^{-1} Å^{-1})	${}^2K_{\phi b}$ (kcal mol^{-1} Å^{-1})	${}^3K_{\phi b}$ (kcal mol^{-1} Å^{-1})
C—C/H—C—C—H	−52.239	−0.784	−0.901
C—C/H—C—C—C	−49.111	−3.173	−0.393
C—C/C—C—C—C	−47.281	−5.641	0.315

K. Bond/Torsion (type 2): $E = (b' - b_0')[{}^1K_{\phi b'}\cos\phi + {}^2K_{\phi b'}\cos 2\phi + {}^3K_{\phi b'}\cos 3\phi]$

Bond/Torsion	${}^1K_{\phi b'}$ (kcal mol^{-1} Å^{-1})	${}^2K_{\phi b'}$ (kcal mol^{-1} Å^{-1})	${}^3K_{\phi b'}$ (kcal mol^{-1} Å^{-1})
H—C/H—C—C—H	0.778	0.476	0.079
H—C/H—C—C—C	0.779	0.288	0.398
C—C/H—C—C—C	1.665	0.583	−0.342
C—C/C—C—C—C	2.294	0.765	0.357

L. Angle/Torsion: $E = (\theta - \theta_0)[{}^1K_{\phi\theta}\cos\phi + {}^2K_{\phi\theta}\cos 2\phi + {}^3K_{\phi\theta}\cos 3\phi]$

Angle/Torsion	${}^1K_{\phi\theta}$ (kcal mol^{-1} rad^{-1})	${}^2K_{\phi\theta}$ (kcal mol^{-1} rad^{-1})	${}^3K_{\phi\theta}$ (kcal mol^{-1} rad^{-1})
H—C—C/H—C—C—H	−1.622	0.739	−0.449
H—C—C/H—C—C—C	0.260	0.702	−0.163
C—C—C/H—C—C—C	−1.886	0.101	−0.295
C—C—C/C—C—C—C	−0.030	−0.015	−0.003

whole, the largest errors in the energy and in the first and second derivatives are found for methylcyclopropane. The average first derivative deviations for all molecules are 10.1 and 33.0% for QMFF and the harmonic force fields, while the average second derivative deviations are 4.6 and 18.4% for the same two force fields. Thus, neglecting anharmonicity and cross-term interactions causes an increase in the error in calculated forces and curvatures by factors of 3.3 and 4.0, respectively.

EFFECT OF STRAIN ON QUALITY OF FIT

These results dramatically emphasize the large degree of anharmonicity and intramolecular coupling interactions that exist among even the alkane

TABLE III.
Ability of Diagonal Harmonic (Class I) Force Field and the Class II Functional Form to Account for the Quantum Mechanical Energy Surface of Alkanes: rms Deviations from *Ab Initio* Energies and First and Second Derivatives of Distorted Structures of Alkanes for a Harmonic Diagonal Force Field and for QMFF.

	ΔE_{rms}[a] (kcal/mol)		$\%\Delta E_{rms}$[b]		$\%\Delta E'_{rms}$[c]		$\%\Delta E''_{rms}$[d]	
Molecule	Harmonic	QMFF	Harmonic	QMFF	Harmonic	QMFF	Harmonic	QMFF
Methane	1.09	0.26	5.3	1.3	15.2	7.6	16.2	3.9
Ethane	1.24	0.40	3.9	1.3	26.8	7.7	12.8	2.6
Propane	2.47	1.54	4.5	2.8	26.0	13.5	14.3	3.6
n-Butane	2.55	0.56	3.8	0.8	28.6	6.2	22.0	3.2
n-Pentane	0.37	0.64	1.3	2.3	26.8	10.2	14.1	2.9
Isobutane	1.78	0.62	2.6	0.9	26.5	16.1	18.1	4.7
Isopentane	0.83	0.73	4.0	3.5	25.7	6.7	15.9	3.3
Neopentane	1.40	0.46	6.5	2.1	29.5	7.1	17.1	4.1
Cyclopentane	2.85	1.61	4.1	2.3	30.0	10.5	18.9	4.7
Cyclohexane	1.60	0.74	4.5	2.1	27.2	8.2	13.1	2.9
Cyclobutane	3.71	0.43	17.0	2.0	19.5	7.0	19.6	4.9
Methylcyclobutane	5.60	0.78	9.1	1.3	25.9	8.2	24.9	5.0
Cyclopropane	2.39	0.34	11.2	1.6	65.7	12.4	22.7	6.2
Methylcyclopropane	5.96	2.87	14.3	6.9	64.8	19.5	19.3	7.4
1,2-Dimethylcyclopropane	2.85	0.86	5.4	1.6	56.0	11.6	21.5	7.1
1,1-Dimethylcyclopropane	6.33	1.09	13.1	2.3	43.5	9.7	24.5	7.3
Average	2.69	0.87	6.9	2.2	33.0	10.1	18.4	4.6

[a]rms deviation of force-field relative energies from *ab initio* relative energies.
[b]Percent (rms) deviation of force-field relative energies, calculated by dividing ΔE_{rms} by the maximum relative energy of distortion (listed in Table I).
[c]Percent (rms) deviation of force-field first derivatives of the energy.
[d]Percent (rms) deviation of force-field second derivatives of the energy.

molecules, and in the alkane energy surface. The inclusion of terms to represent this anharmonicity and the coupling interactions greatly increases the ability of the analytic form to account for the energy surface of these alkanes and the physical and dynamic properties that we wish to calculate. It becomes clear that this allows us to achieve a significant improvement in our ability to simulate molecular properties of interest as they all derive from the molecular energy surface.

Use of the Modeled Energy Surface to Assess Importance of Individual Terms

It is of interest to assess the importance of individual coupling interactions and of anharmonicity in determining the overall shape of the quantum energy surface. This can be done quantitatively by making use of the analytic energy expression (force field) derived from this surface.

The importance of the individual energy function terms in eqs. (2)–(15) are assessed by the degree to which these terms contribute to the fit of the relative energies and the first and second derivatives of the energy in the sum of squared deviations in eq. (21) (Table IV). This assessment is based on the extent of increases in deviations from *ab initio* energies and energy derivatives as selected terms in the energy function are omitted. The second column in Table IV shows the reduction in the number of adjustable parameters that accompanies leaving out the interactions listed in the first column. The third column presents the rms deviations in relative energies for all 16 molecules used for the fit. Similarly, the fourth and fifth columns are the relative rms deviations in first and second derivatives, respectively. In all cases, the relative deviations were computed from the ratio of squared deviations to squared energy or derivative values.

The importance of individual terms varies widely between differing molecules, with specific terms such as cross-terms often being important. Thus, for example, the angle-angle-torsion cross-term, which is neglected in other force fields, was shown by our analysis of the quantum mechanical energy surface to reduce the maximum error in the

TABLE IV.
Percent rms Deviations[a] of Derivatives[b] Calculated with Various Terms Omitted from the QMFF Functional Form.

Missing Energy Terms	Reduction of Parameters	ΔE[c]	$\%\Delta E'$	$\%\Delta E''$
None	0	0.96	10.0	4.2
Anharmonic bonds	4	1.36	22.7	18.2
Anharmonic angles	6	0.95	10.5	4.8
Bond/bond	3	0.96	10.1	4.5
Bond/angle	4	1.11	21.4	12.3
Angle/angle	6	1.17	12.0	4.3
Torsion	3	3.50	10.2	4.2
Nonbonds	2	3.20	11.5	4.2
Angle/angle/torsion	3	1.85	16.8	6.4
Bond/torsion 1	9	1.69	30.7	4.1
Bond/torsion 2	12	0.93	11.4	4.4
Angle/torsion	12	1.17	11.7	4.4

[a]rms deviations averaged for all 16 molecules in the fit.
[b]E' = first derivative and E'' = second derivatives; $\%\Delta$ is the percent deviation.
[c]Absolute rms energy deviations in kcal/mol.

sampled energies from 5.3 kcal/mol to 1.3 kcal/mol in isobutane.

In the following, we will analyze the importance of individual terms in the force field both as revealed by their contributions to the first and second derivatives of the energy.

FIRST DERIVATIVES

An examination of the first derivative deviations in column 4 of Table IV reveals the terms that make the largest contributions to intramolecular forces in alkanes. The most important term for reducing the deviations in forces is the bond/torsion 1 term [see eq. (14) and Fig. 6f]. An increase from 10.0 to 30.7% in the first derivative deviations results when this term is neglected in the functional form. The bond/torsion coupling interaction is used to model the torsion angle dependence of forces on the bond length, as discussed above [see eq. (16)]. The importance of this coupling term for calculations of bond length forces is consistent with the large increases (0.014 and 0.019 Å) in bond lengths that are observed in the eclipsed ethane and neopentane structures. Table IV shows that the anharmonic bond, bond/angle, and angle/angle/torsion terms are the three other most important terms for reproducing intramolecular forces in alkanes. When these terms are neglected, the first derivative deviations increase from 10.0 to 22.7, 21.4, and 16.8%, respectively. This indicates that these three terms are also important for calculations of structural changes in strained alkanes.

SECOND DERIVATIVES

The most important cross-terms for fitting the curvature of the energy surface are the bond/angle and angle/angle/torsion terms. From the last column in Table IV, it can be seen that neglecting these two cross-terms causes the second derivative deviations to increase from 4.2 to 12.3 and 6.4%, respectively. For fitting *ab initio* second derivatives, these two cross-terms are evidently less important than bond anharmonicity but much more important than bond angle anharmonicity or other coupling interactions. This suggests that these two cross-term types should be particularly important for accurate calculations of vibrational frequencies.

BOND ANHARMONICITY

The importance of anharmonicity in the bond energy function is indicated by the results in the second row of Table IV. These results were obtained by omitting the cubic and quartic bond energy terms from eq. (2) and calculating the energy and derivatives, along with the rms deviations from the *ab initio* results, with the remaining QMFF energy function terms and parameters in Table II. The results in Table I show that elimination of bond anharmonicity costs 0.4 kcal/mol in the fit of *ab initio* energies, that is, the rms energy deviation increased from 0.96 to 1.36 kcal/mol. Errors in the first derivatives more than double (from 10.0 to 22.7%) when bond anharmonicity is neglected, while errors in the second derivatives increase from

4.2 to 18.2%. This fourfold increase in the second derivative error dramatically demonstrates the importance of bond anharmonicity in the *ab initio* potential energy surface.

Conclusions

A general method for determining force fields for potential energy functions from quantum mechanical energy surfaces has been presented. The functional form of the resulting force field, a Class II force field, was given and the method was used to derive an alkane quantum mechanical force field, that is, a force field that fits the quantum mechanical surface. All force constants and reference values were determined. The force field was tested in terms of the fit to the *ab initio* energy surface by comparing energies and first and second derivatives. The results showed that the quality of the fit is generally good. The quantum mechanical energy surface was found to contain significant anharmonicity and intramolecular coupling interactions. This was demonstrated by comparing the Class II force field derived here including analytic representations of these effects with a diagonal quadratic Class I force field, that is, the functional form used commonly in potential functions such as AMBER and CHARMM for the simulation of organic and biomolecular systems. (It was pointed out that CVFF, which contains some coupling terms, is somewhere between a Class I and Class II force field.)

Finally, the importance of the various energy terms was determined quantitatively by assessing the effect on the fit to the energy, first, and second derivatives from omitting the individual terms. In a subsequent article, the quantum force field derived here will be scaled to a fit of experimental data and an experimental force field will be derived. The quantum force field presents enough information on the shape and relative values of the force constants such that only a limited number of experimental data need to be used and only a few scaling factors and reference values need to be determined, thus greatly reducing the amount of information required from experimental data.

Finally, it was pointed out that this methodology should provide a useful starting point for use in simulations for functional groups where insufficient amounts of experimental data are available or where, for pragmatic reasons, there is a need to proceed without fitting experimental data. The protocol described here can be used to derive a force field for any functional group for which energies and first and second derivatives can be obtained from quantum mechanics. The result will be an analytic representation that, if used in a simulation, will give essentially comparable results to that which would have been obtained had one used the quantum energy surface itself to within the range of deviations reported here.

Acknowledgment

The authors gratefully acknowledge the support of the Consortium for Research and Development of Potential Energy Functions. Former and present members of the Consortium include Abbott Laboratories; Battelle Molecular Science Research Center; Bayer AG; Bristol-Myers Squibb Pharmaceutical Group; Convex Computer Corp.; Cray Research, Inc.; Deutsches Kunstoff Institute; Dow Chemical Co.; DuPont Merck Pharmaceutical Co.; E. I. DuPont de Nemours & Co.; Eastman Kodak Co.; Esteve SA, Laboratorios de Dr; ETA Systems; Farmitalia Carlo Erba; Fujitsu Systems Business of America; Gessellschaft für Biotechnologische Forschung mbH; Glaxo Group Research Ltd.; Hitachi, Ltd.; Hybritech, Inc.; IBM Almaden Research Center; ICI Americas, Inc.; Immunex Research & Development Corp.; Institut Francais du Pétrole; Janssen Research Foundation; Ludwig Institute; Mitsubishi Petrochemical; Marion Merrell Dow Research Institute; Merck, Sharp & Dohme Research; Monsanto Co.; National Cancer Institute; North Carolina Supercomputing Center; Protein Engineering Research Institute; Rhône-Poulenc; Rohm & Haas Co.; Sandia National Laboratories; Sandoz, Ltd.; Searle Research & Development; Shell Research, Ltd.; SINTEF; SRI Corp.; Synergie Technologies, Ltd.; Takeda Chemical Industries, Ltd.; Universite De Lille I Cerim; Universite De Nancy I; U.S. Army, Ballistic Research Laboratory and Virology Division; and Upjohn Laboratories.

References

1. A. T. Brunger, J. Kuriyan, and M. Karplus, *Science*, **235**, 458 (1987).
2. G. M. Clore, M. Nilges, D. K. Sukumaran, A. T. Brunger, M. Karplus, and A. M. Gronenborn, *EMBO J.*, **5**, 2729 (1986).
3. P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman, *Science*, **235**, 574 (1987).

4. A. Warshel and F. Sussman, *Proc. Natl. Acad. Sci. USA*, **83,** 3806 (1986).
5. F. Avbelj, J. Moult, D. H. Kitson, M. N. G. James, and A. T. Hagler, *Biochemistry*, **29,** 8658 (1990).
6. M.-J. Hwang, T. P. Stockfisch, and A. T. Hagler, submitted.
7. L. S. Bartell, *J. Chem. Phys.*, **32,** 827 (1960); L. S. Bartell, *Tetrahedron*, **17,** 177 (1962).
8. S. Fitzwater and L. S. Bartell, *J. Am. Chem. Soc.*, **98,** 5107 (1976).
9. J. B. Hendrickson, *J. Am. Chem. Soc.*, **83,** 4537 (1961); J. B. Hendrickson, *J. Am. Chem. Soc.*, **86,** 4854 (1964).
10. T. Simanouti, *J. Chem. Phys.*, **17,** 734 (1949); T. Simanouti and S. Mizushima, *J. Chem. Phys.*, **17,** 1102 (1949).
11. J. H. Schachtschneider and R. G. Snyder, *Spectrochim. Acta*, **19,** 117 (1963); R. G. Snyder and J. H. Schachtschneider, *Spectrochim. Acta*, **21,** 169 (1965).
12. S. Lifson and A. Warshel, *J. Chem. Phys.*, **49,** 5116 (1968).
13. A. Warshel and S. Lifson, *J. Chem. Phys.*, **53,** 582 (1970).
14. A. T. Hagler, P. S. Stern, S. Lifson, and S. Ariel, *J. Am. Chem. Soc.*, **101,** 813 (1979).
15. S. Lifson and P. S. Stern, *J. Chem. Phys.*, **77,** 4542 (1982).
16. R. H. Boyd, *J. Chem. Phys.*, **49,** 2574 (1968).
17. C. Shieh, D. McNally, and R. H. Boyd, *Tetrahedron*, **25,** 3653 (1969).
18. S. Chang, D. McNally, S. Shary-Tehrany, S. M. J. Hickey, and R. H. Boyd, *J. Am. Chem. Soc.*, **92,** 3109 (1970).
19. N. L. Allinger, M. A. Miller, F. A. Van Catledge, and J. A. Hirsch, *J. Am. Chem. Soc.*, **89,** 4345 (1967); N. L. Allinger, J. A. Hirsch, M. A. Miller, I. J. Tyminski, and F. A. Van Catledge, *J. Am. Chem. Soc.*, **90,** 1199 (1968).
20. N. L. Allinger, *J. Am. Chem. Soc.*, **99,** 8127 (1977).
21. N. L. Allinger, Y. H. Yuh, and J.-H. Lii, *J. Am. Chem. Soc.*, **111,** 8551 (1989).
22. J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, **111,** 8566 (1989).
23. J.-H. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, **111,** 8576 (1989).
24. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.*, **7,** 230 (1986).
25. L. Nilsson and M. Karplus, *J. Comp. Chem.*, **7,** 591 (1986).
26. J. R. Maple, U. Dinur, and A. T. Hagler, *Proc. Natl. Acad. Sci. USA*, **69,** 5350 (1988).
27. A. T. Hagler, J. R. Maple, T. S. Thacher, G. B. Fitzgerald, and U. Dinur, In *Computer Simulation of Biomolecular Systems—Theoretical and Experimental Applications*, van Gunsteren and Weiner, Eds., ESCOM, Leiden, 1989, pp. 149–167.
28. U. Dinur and A. T. Hagler, In *Reviews in Computational Chemistry*, Vol. 2, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, chap. 4, pp. 99–164.
29. A. T. Hagler, In *The Peptides*, Vol. 7, V. J. Hruby and J. Meienhofer, Eds., Academic, New York, 1985, pp. 213–299.
30. S. Dasgupta and W. A. Goddard, *J. Chem. Phys.*, **90,** 7207 (1989).
31. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York, 1986.
32. C. E. Blom and C. Altona, *Mol. Phys.*, **31,** 1377 (1976).
33. M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, L. R. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol, and J. A. Pople, Gaussian 88, Gaussian, Inc., Pittsburgh, PA, 1988.
34. E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, Dover Publications, New York, 1955.
35. J. P. Bowen and N. L. Allinger, In *Reviews in Computational Chemistry*, Vol. 2, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, chap. 3, pp. 81–97.
36. P. M. Morse, *Phys. Rev.*, **34,** 57 (1929).
37. U. Burkert and N. L. Allinger, *Molecular Mechanics*, American Chemical Society, Washington, DC, 1982.
38. M. Waldman and A. T. Hagler, *J. Comp. Chem.*, **14,** 1077 (1993).
39. T. Oie, G. M. Maggiora, R. E. Christoffersen, and D. J. Duchamp, *Int. J. Quantum Chem. Quantum Biol. Symp.*, **8,** 1 (1981).
40. T. A. Halgren, *J. Am. Chem. Soc.*, **112,** 4710 (1990).
41. A. Warshel and S. Lifson, *Chem. Phys. Lett.*, **4,** 255 (1969).
42. L. Norskov-Lauritsen and N. L. Allinger, *J. Comp. Chem.*, **5,** 326 (1984).
43. R. Fletcher, *Practical Methods of Optimization*, Vol. 1, Wiley, New York, 1980.
44. N. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed., Wiley, New York, 1981.