Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules

M. J. Hwang, T. P. Stockfisch, and A. T. Hagler*

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

Received July 19, 1993. Revised Manuscript Received December 21, 1993[®]

Abstract: A second generation Class II force field is derived for the alkyl group and alkane molecules. The Class II functional form is presented and force constants are given. The criteria that define this second generation force field are the following: (1) it accounts for the properties of both isolated small molecules, condensed phases, and macromolecular systems and (2) the functional form is characterized by being anharmonic, with quartic stretching and quartic angle bending, and includes a variety of important intramolecular coupling interactions. It is also characterized by a soft repulsion, either 9th power or exponential, rather than the more usual 12th power repulsion. The force field is derived by scaling an analytical representation derived from a fit to a quantum mechanical energy surface. Only seven parameters were needed in this scaling to reproduce 150 experimental observables. The resulting Class II force field is shown to fit the structural, energetic, and dynamic properties of the alkane molecules comprising the training set well. These molecules include small acyclic alkanes, strained molecules such as isobutane, and small rings including cyclopropane and cyclobutane. Thus the properties of highly strained molecules including small rings are accounted for with one set of transferable parameters. The results are compared with those obtained from Class I diagonal quadratic force fields commonly used in simulations of biological systems and the Class II functional form is shown to reproduce trends unattainable by the simpler forms where anharmonicity and coupling interactions are not accounted for. Most dramatic is its ability to fit the small ring compounds, cyclopropane and cyclobutane, with the same transferable energy functions that account for larger rings and acyclic molecules. This is the first energy surface able to achieve this range of applicability, a degree of transferability hypothesized in the literature to be unachievable. Finally, and perhaps most importantly it is pointed out that the methodology presented here provides a paradigm for the straightforward derivation of force fields for arbitrary molecules of interest even where experimental data is sparse or missing. The force field can be derived based on the techniques described here as long as the quantum mechanical calculation can be carried out. The resulting quantum force field, at that point, could either be used on its own or scaled to provide a pragmatic and reasonably accurate force field.

I. Introduction

Force fields, or analytical approximations to molecular energies as functions of atomic coordinates, are becoming increasingly employed in a variety of molecular mechanics and dynamics applications requiring accurate simulation of complex systems. Hence, there is a pressing need for force fields of increased accuracy and applicability to general classes of organic, biomolecular, polymeric, and bioinorganic systems.

In a preceding paper we introduced a methodology for deriving a *quantum mechanical force field* (QMFF) and applied it to alkanes.[1] That is, we obtained an analytical expression which reproduced the Hartree–Fock (HF) energy surface of a wide variety of alkane molecules. The method involved makes use of quantum mechanical relative energies and the first and second Cartesian derivatives of the energy for a set of distorted alkane molecules as the "observables".[2] The force field was then derived by fitting the corresponding quantities calculated from the analytical energy expression to this large set of data describing the quantum energy surface, in much the same way that force fields have previously been derived by fitting experimentally observed properties such as structures, energies, and frequencies.[3–13]

There are many significant advantages to the use of quantum mechanics. The number of "observables" that may be included in the fit is in principle unlimited, and they may be derived for any functional group for which a quantum mechanical calculation can be carried out. If a particular force constant or coupling term in the force field is underdetermined, additional information in the form of these "quantum observables" may be obtained easily. This can be accomplished simply by calculating the energies and derivatives of several additional distorted configurations of a variety of molecules in which the bond, angle, or other internal coordinate is sampled over a range of values.[1] Thus, the quantum approach yields a wealth of information about molecular energy surfaces. As outlined in this study[1], this provides enough information to allow determination of not only the harmonic diagonal constants (Class I) as used in most standard force fields for structural biology[12–14] but also the important anharmonic and coupling interactions.[1,2,5,6,15,16] At the same time, the quantum mechanical surfaces demonstrate the importance

---

[®] Abstract published in *Advance ACS Abstracts*, February 15, 1994.

(1) Maple, J. R.; Hwang, M.-J.; Stockfisch, T.; Dinur, U.; Waldman, M.; Ewig, C.; Hagler, A. T. *J. Comp. Chem.* In press.

(2) Maple, J. R.; Dinur, U.; Hagler, A. T. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *69*, 5350–5353.

(3) Fitzwater, S.; Bartell, L. S. *J. Am. Chem. Soc.* **1976**, *98*, 5107–5115.

(4) Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582–594.

(5) Hagler, A. T.; Stern, P. S.; Lifson, S.; Ariel, S. *J. Am. Chem. Soc.* **1979**, *101*, 813–819.

(6) Lifson, S.; Stern, P. S. *J. Chem. Phys.* **1982**, *77*, 4542–4550.

(7) Boyd, R. H. *J. Chem. Phys.* **1968**, *49*, 2574–2583.

(8) Chang, S.; McNally, D.; Shary-Tehrany, S.; Hickey, S. M.; Boyd, R. H. *J. Am. Chem. Soc.* **1970**, *92*, 3109–3118.

(9) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

(10) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566–8575.

(11) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576–8582.

(12) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230–252.

(13) Nilsson, L.; Karplus, M. *J. Comp. Chem.* **1986**, *7*, 591–616.

(14) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. *Proteins: Struct. Funct. Genet.* **1988**, *4*, 31–47.

(15) Dinur, U.; Hagler, A. T. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, Chapter 4, pp 99–164.

of these additional interactions.[1,2,15,16] Thus, it was shown that one could use quantum mechanics to determine the detailed functional form, i.e. to reveal the importance of various quantities such as anharmonicity and cross terms which have been the subject of discussion and which have been difficult to determine in the past—generally because of a dearth of experimental data. The importance of these terms in reproducing experimental trends will be elucidated below by comparing properties calculated with the standard Class I and new Class II energy functions derived here.

Another advantage of using ab initio observables, i.e., the energy and its first and second derivatives, is that there is a one-to-one relationship between the calculated and observed properties. That is the energy and the first and second derivatives are both the "quantum observables" and the calculated properties. No model needs to be interposed between the calculated and observed properties. This is not the case when fitting experimental properties. For example, when comparing observed and calculated vibrational frequencies a harmonic model is usually invoked in order to calculate the frequencies from the potential function. In this case a mathematical model of frequencies—the harmonic model—is invoked to obtain the calculated quantity (while the observed frequencies contain anharmonic and other effects) and there is not a unique one-to-one correspondence between "calculated" and "observed" quantities. In fitting "quantum observables" the second derivatives themselves are the "observed" properties, rather than the derived observed vibrational frequency which is related to it indirectly. Thus, the derivation of a force field from quantum observables is more direct and, when there are deviations or other inadequacies in the derived force field, these are more easily analyzed and understood.

The ability of the QMFF to reproduce the quantum mechanical energy surface was assessed by its ability to fit the configurational energies and first and second derivatives of the distorted molecules[1] and tested by comparing the more standard physical observables (equilibrium structures, vibrational frequencies, conformational energies, and rotational barriers) calculated from the derived force field with the corresponding quantities *calculated from the quantum hamiltonian.*[17] In this way it was shown that the functional form derived by this procedure accurately describes the quantum mechanical energy surface.[1,2] The force field determined in this way is given in eq 1. Here b, θ, ϕ, and χ are bond, bond angle, torsion angle, and out-of-plane internal coordinates (note that the out-of-plane internal does not occur in alkanes; it has been included in this expression for completeness; for a discussion of this term, see ref 2), respectively, with the subscript 0 denoting a reference value, and ^2K_b, $^2K_\theta$, etc. are the force constants determined from *the quantum mechanical energy surface.* Terms involving explicit internuclear distances, r, represent nonbond interactions.

Transferability. In the present work dealing with saturated hydrocarbons, force constants are totally transferable. That is, they depend on whether the bonded atoms in the internal coordinates are carbon or hydrogen, but they do not depend on any other details of the substructural moieties in which they are found. Thus, they are independent of whether carbon is primary, secondary, or tertiary, or belonging to a large or small ring. This property is extremely important if our ultimate purpose is to derive a general force field which we wish to use to extrapolate to molecules which are not part of the training set and may be in environments not sampled in the training set. It is perhaps worth taking a moment to elaborate on this point.

There are two reasons this is important. The first reason is that if we assume for the moment that this characteristic indeed

(16) Hagler, A. T.; Maple, J. R.; Thacher, T. S.; Fitzgerald, G. B.; Dinur, U. In *Computer Simulation of Biomolecular Systems—Theoretical and Experimental Applications*; van Gunsteren, Weiner, Eds.; ESCOM: Leiden, 1989; pp 149–167.

(17) Maple, J. R.; Hwang, M.-J.; Stockfisch, T. P.; Hagler, A. T., to be submitted for publication.

$$E = \sum_b [^2K_b(b - b_0)^2 + {}^3K_b(b - b_0)^3 + {}^4K_b(b - b_0)^4]$$

$$+ \sum_\theta [^2K_\theta(\theta - \theta_0)^2 + {}^3K_\theta(\theta - \theta_0)^3 + {}^4K_\theta(\theta - \theta_0)^4]$$

$$+ \sum_\phi [^1K_\phi(1 - \cos \phi) + {}^2K_\phi(1 - \cos 2\phi) + {}^3K_\phi(1 - \cos 3\phi)]$$

$$+ \sum_\chi K_\chi \chi^2 + \sum_{i>j} \frac{q_i q_j}{r_{ij}} + \sum_{i>j} \epsilon \left[2\left(\frac{r^*}{r_{ij}}\right)^9 - 3\left(\frac{r^*}{r_{ij}}\right)^6 \right]$$

$$+ \sum_b \sum_{b'} K_{bb'}(b - b_0)(b' - b'_0) + \sum_\theta \sum_{\theta'} K_{\theta\theta'}(\theta - \theta_0) \times$$

$$(\theta' - \theta'_0)$$

$$+ \sum_b \sum_\theta K_{b\theta}(b - b_0)(\theta - \theta_0)$$

$$+ \sum_\phi \sum_b (b - b_0)[^1K_{\phi b} \cos \phi + {}^2K_{\phi b} \cos 2\phi + {}^3K_{\phi b}\cos 3\phi]$$

$$+ \sum_\phi \sum_{b'} (b' - b'_0)[^1K_{\phi b'} \cos \phi + {}^2K_{\phi b'} \cos 2\phi +$$

$$^3K_{\phi b'} \cos 3\phi]$$

$$+ \sum_\phi \sum_\theta (\theta - \theta_0)[^1K_{\phi\theta} \cos \phi + {}^2K_{\phi\theta} \cos 2\phi + {}^3K_{\phi\theta} \cos 3\phi]$$

$$+ \sum_\phi \sum_\theta \sum_{\theta'} K_{\phi\theta\theta'} (\theta - \theta_0)(\theta' - \theta'_0) \cos \phi \tag{1}$$

reflects the true energy surface of hydrocarbons, namely the intrinsic force constants are independent of the substructural moiety in which they are found and the environmental effects on observed properties such as lengthening of bonds or differing angles in primary and secondary carbons are due to neighboring effects, then the fact that we obtain a transferable force field reflects that it is more accurately representing the true energy surface. If the functional form is deficient, then what are truly environmental effects which couple to and distort the bond or other internals would need to be represented by using different force constants for the same internal. That is, let us say a C–C bond couples to a second C–H or C–C bond differently in primary or secondary environments, which lead to different "unperturbed" bond lengths. Then what is really bond–bond coupling would be interpreted in a diagonal force field (no coupling) *as different intrinsic force constants* for primary and secondary C–C bonds. However, these effective force constants *would then not be transferable* to different geometric environments of the given bond since the couplings which determine the differences are not included in the analytical representation. We will see examples of this below.

The second reason that this is important is that the fewer the types of force constants and the more transferable they are, the better they will be able to achieve the ultimate objective of the force field which is to calculate properties of systems which have not been addressed previously. As we shall see, based on the results shown below, one of the characteristics of the Class II force field is that it is better able to handle trends in structural properties of the hydrocarbons and other molecules when the environment is changed through a change in conformation or substituent. This again tends to confirm the hypothesis that the molecular energy surface is more accurately represented by a functional form which accounts for anharmonicity and coupling interactions.

In this paper, we take the derivation of the Class II energy surface a step further to produce a force field (from the QMFF) that accounts for, and predicts, the *experimental* properties of the hydrocarbon functional group and alkane molecules. To refine the QM force field we now return to the same set of experimental observables used in traditional techniques of force field derivation including gas-phase structures, vibrational frequencies, rotational

barriers, and conformational energies. The difference between this and previous approaches[3–14] is that rather than attempting to determine the entire force field from limited experimental data, we use only a small set of scale factors and reference values to correct the quantum mechanical energy surface for well-known and systematic deviations in Hartree–Fock calculations.

The scaling of the QM force field is desirable because although the quantum mechanical calculations have the advantages outlined above, and have been shown to account well for molecular structures and conformational energies,[18] calculations at the Hartree–Fock level have deficiencies. It is known that force constants derived at this level are roughly 10–15% too large.[18] (There are smaller systematic errors in bond lengths also.) Thus, in previous applications where force constants have been derived directly from the Hartree–Fock Hessians, they have been scaled[19–29] to fit observed vibrational frequencies. Another deficiency is that Hartree–Fock level calculations do not include dispersion interactions. However, nonbond parameters can be determined by fitting experimental intermolecular properties, especially crystal structures and lattice energies.[30–34] Following a common pragmatic practice,[6,35–37] the derived nonbond parameters were kept fixed during the parametrization of intramolecular force constants in order to remove the correlation between them.[1] We have found this to be a satisfactory approach for a variety of functional groups, not requiring additional iterations to further refine the parameters. Derivation of the nonbond parameters from crystal data as well as validation on other condensed-phase properties will be reported elsewhere. In summary, at this stage we scale the intramolecular force field we obtained from the quantum energy surface (QMCFF93)[1] to fit a wide range of experimental data in order to obtain the desired Class II energy surface.

As noted above the reason for the scaling is because of limitations in the quantum methods. There are higher level methods available at this time and the question arises (also raised by one of the referees) as to what the effect might be if higher level calculations were carried out, for example, those containing electron correlation or polarization functions on hydrogens. The choice of HF/6-31G* was made for pragmatic reasons as it gives energies and geometries in reasonable agreement with higher level calculations. Thus, where the results have been checked, no significant differences have been found for the alkanes. Clearly systematic study is warranted here and it should be pointed out

that as the quantum methodology improves and becomes computationally more accessible either through algorithmic changes or new methodologies such as density functional theory, these advances can be incorporated directly into the methodology proposed here. At some point, as the quantum techniques approach experimental accuracy, one would hope that no scaling would be required and one could obtain an analytical representational of the quantum energy surface with no further scaling that fit the experimental molecular properties to within experimental accuracy. Clearly at this point, the impact of any errors in 6-31G* in the level of the Hartree–Fock and the limitation of the method is in the degree to which the ratio of different force constants is in error. Thus if the error in the cubic term is different than the error in the quadratic term or if the error in a C–C–C bond angle is different than the error in a C–C–N bond angle, and they are not correlated within the quantum method, then the resulting force field will be off. The ultimate test of the adequacy of the method is of course in how well the experimental data is fit. Based on the results to date, the approximation seems to be reasonable.

In fact, as pointed out in ref 1, the methodology proposed here can be considered a logical extension of Lifson's consistent force field method for deriving a force field from as wide a variety of experimental data as possible. The difference is that it first takes advantage of the vast amount of data available from quantum mechanics. However, it then uses the same wide variety of experimentally available data at the heart of the CFF methodology developed by Lifson.[4–6]

Paradigm for Force Field Derivation. The methodology described here provides us with a paradigm for force field development, where our ultimate goal is to determine force fields for a wide range of molecules often involving functional groups where there is little or no experimental data upon which to base a rigorous force field. That is, faced with a new functional group in a system where force field parameters do not exist and experimental data are sparse, the methodology outlined above can be invoked. Either the QMFF energy functions and constants can be used directly yielding a molecular mechanics or dynamics result of comparable quality to that which would result if carried out on the quantum energy surface, or scaling factors can be transferred. The method for determining this small group of scaling parameters is described below in the Methods section.

The combined use of quantum mechanics and experimental data in this way allows us to significantly improve the description of the molecular energy surface. We have defined a set of criteria in Table 1 for the enhanced "Class II" force field, which at the same time serve to define it and distinguish it from previous generation "Class I" force fields which are, for the most part, diagonal, quadratic representations. We will compare results obtained from the Class II energy surface with some of the standard Class I representations below.

The organization of this paper is as follows: In Section II the scaling method is described. In Section III we present the entire CFF93 alkane force field and demonstrate how the number of adjustable parameters is drastically reduced with the present method. Calculations and comparison against experimental structural, energetic, and dynamic (vibrational) properties are reported in Section IV. Finally, the results are summarized in Section V.

II. Scaling Method

As noted above, the complete CFF93 force field is obtained by scaling the quantum force field derived from the ab initio energy surface.[1] This is illustrated in eq 2,

(18) Hehre, W. J.; Radom, L.; Schleyer, P.v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York 1986.

(19) Bolm, C. E.; Slingezland, P. J.; Altona, C. *Mol. Phys.* **1976**, *31*, 1359–1376. Blom, C. E.; Altona, C. *Mol. Phys.* **1976**, *31*, 1377–1391.

(20) Fogarasi, G.; Pulay, P. *Vib. Spectra Struct.* **1985**, *14*, 125–219.

(21) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. *J. Am. Chem. Soc.* **1979**, *101*, 2550–2560.

(22) Shaw, R. A.; Ibrahim, N.; Wieser, H. *Can. J. Chem.* **1990**, *68*, 90–101.

(23) Cheam, T. C.; Krimm, S. *J. Mol. Struct.* **1989**, *193*, 1–34.

(24) Dutler, R.; Rauk, A.; Shaw, R. A. *J. Phys. Chem.* **1990**, *94*, 118–124.

(25) Somogyi, A.; Jalsovszky, G.; Fulop, C.; Stark, J.; Boggs, J. E. *Spectrochim. Acta* **1989**, *45A*, 679–693.

(26) Jalkanen, K. J.; Stephens, P. J. *J. Phys. Chem.* **1991**, *95*, 5446–5454.

(27) McKean, D. C.; McQuillan, G. P.; Murphy, W. F.; Zerbetto, F. *J. Phys. Chem.* **1990**, *94*, 4820–4831; **1993**, *97*, 581–595. Murphy, W. F.; Fernandez-Sanchez, J. M.; Raghavachari, K. *J. Phys. Chem.* **1991**, *95*, 1124–1139.

(28) Bock, Ch. W.; McDiarmid, R. *J. Mol. Struct.* **1990**, *222*, 415–429.

(29) Koframek, M.; Karpfen, A.; Lischka, H. *Int. J. Quantum Chem.* **1990**, *24*, 721–732.

(30) Hagler, A. T.; Lifson, S. *Acta Crystallogr.* **1974**, *B30*, 1336–1341.

(31) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319–5327.

(32) Hagler, A. T.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5327–5335.

(33) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111–5121.

(34) Hagler, A. T.; Lifson, S.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5122–5130.

(35) Pietila, L.-O.; Mannfors, B.; Palmo, K. *J. Mol. Struct.* **1990**, *218*, 315–320.

(36) Palmo, K.; Pietila, L.-O.; Krimm, S. *J. Comput. Chem.* **1991**, *12*, 385–390.

(37) Aleman, C.; Canela, E. I.; Franco, R.; Orozco, M. *J. Comput. Chem.* **1991**, *12*, 664–674.

2518 *J. Am. Chem. Soc., Vol. 116, No. 6, 1994*

Hwang et al.

$$E = S_b\{\sum_b [^2K_b(b - b_0)^2 + ^3K_b(b - b_0)^3 + ^4K_b(b - b_0)^4]\}$$

$$+ S_\theta\{\sum_\theta [^2K_\theta(\theta - \theta_0)^2 + ^3K_\theta(\theta - \theta_0)^3 + ^4K_\theta(\theta - \theta_0)^4]\}$$

$$+ S_\phi\{\sum_\phi [^1K_\phi(1 - \cos \phi) + ^2K_\phi(1 - \cos 2\phi) +$$

$$^3K_\phi(1 - \cos 3\phi)]\}$$

$$+ S_\chi\{\sum_\chi K_\chi\chi^2\} + \sum_{i>j} \frac{q_iq_j}{r_{ij}} + \sum_{i>j}\epsilon\left[2\left(\frac{r^*}{r_{ij}}\right)^9 - 3\left(\frac{r^*}{r_{ij}}\right)^6 \right]$$

$$+ S_c\{\sum_b\sum_{b'}K_{bb'}(b - b_0)(b' - b'_0) + \sum_\theta\sum_{\theta'}K_{\theta\theta'} \times$$

$$(\theta - \theta_0)(\theta' - \theta'_0)$$

$$+ \sum_b\sum_\theta K_{b\theta}(b - b_0)(\theta - \theta_0)$$

$$+ \sum_\phi\sum_b (b - b_0)[^1K_{\phi b} \cos \phi + ^2K_{\phi b} \cos 2\phi + ^3K_{\phi b} \cos 3\phi]$$

$$+ \sum_\phi\sum_{b'} (b' - b'_0)[^1K_{\phi b'} \cos \phi + ^2K_{\phi b'} \cos 2\phi +$$

$$^3K_{\phi b'}\cos 3\phi]$$

$$+ \sum_\phi\sum_\theta (\theta - \theta_0)[^1K_{\phi\theta} \cos \phi + ^2K_{\phi\theta} \cos 2\phi + ^3K_{\phi\theta} \cos 3\phi]$$

$$+ \sum_\phi\sum_\theta\sum_{\theta'}K_{\phi\theta\theta'}(\theta - \theta_0)(\theta' - \theta'_0) \cos \phi\} \qquad (2)$$

where S_b, S_θ, S_ϕ, S_χ, and S_c are scale factors for bond stretching, angle bending, torsion, out-of-plane, and cross term force constants, respectively.

One can immediately see from this equation the reduction in the number of parameters which need to be adjusted to fit the experimental data achieved by this technique. Rather than needing to parametrize each of the force constants individually by fitting the experimental observables, as done traditionally, we need only determine the scale factors for each class of internal deformation. Thus, the angle, torsion, and out-of-plane terms each have one scaling factor associated with them for all types of angles, torsions, etc., while individual scaling factors were used for C–C and C–H bonds. All cross terms are scaled by the same factor. In addition to force constant scale factors we also make slight adjustments to bond reference values, b_0, since bond lengths calculated in the Hartree–Fock approximation are generally too short. In alkanes we have two bond reference values, C–C and C–H, and we use two corresponding bond reference adjustments. Further work on different functional groups has shown an additional sought after advantage in that the scale factors may be transferred reasonably well. This is a property of clear practical as well as theoretical importance since, if it holds up, it implies that reasonably accurate Class II force fields may be obtained for any functional group amenable to the Hartree–Fock procedure. The quantum energy surface derived from the ab initio energy surface could then be scaled directly by the set of "transferable" scale factors.

Calculation of Scale Factors and Reference Values. Since there are only four types of scale factors to be derived in the alkane force field, that is the bond, angle, torsion, and cross term (there are no out of planes), we chose to use a simple, pragmatic approach to determine these, which turned out to be adequate. We determined the scale factors only from the frequency data and thus left the structural and energetic data to provide still additional tests which are in a sense outside the training set for the force field. The Hartree–Fock vibrational frequencies were calculated from the mass weighted Cartesian second derivatives matrix using

Table 1. Characteristics of Class II Force Field

I. The <u>Same</u> force field fits properties of:

　Isolated small molecules (including structural, thermodynamic, spectroscopic data and dipole moments)

　Condensed phases (including crystal lattice vectors, position and orientation of asymmetric units, sublimation energies, liquid heats of vaporization)

　Macromolecular systems

II. <u>Fits</u> highly strained molecules including small rings with the same parameters

III. Analytical Form[a]

　An anharmonic force field characterized by Morse or quartic bond stretching, quartic angle bending

　Well characterized, 1-, 2-, and 3-fold torsion terms

　Cross terms

　Exponential or 9th power nonbonded repulsion

[a] The earliest Class II analytical form resulted from the pioneering work of Lifson,[4-6] who insisted some 20 years ago that the force field should fit a wide range of experimental properties. In addition, the recent MM3 functional form[9-11] is also a Class II functional form. Finally, in the early work of Boyd,[7,8] the importance of coupling was also realized and he introduced a Urey Bradley type function to account for this.

(A) Acyclic Compounds

methane　　ethane　　n-butane　　isobutane

(B) Cyclic Compounds

cyclopropane　　cyclobutane　　cyclohexane
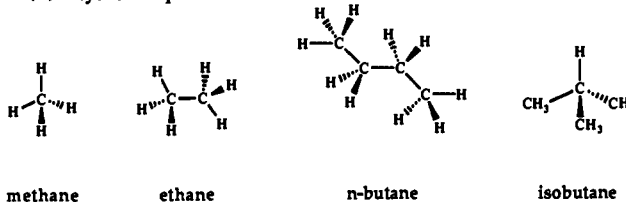
Figure 1.

the standard harmonic approximation.[38] In order to determine the scale factor for the C–H bond, each Hartree–Fock C–H stretching frequency was compared to the experimental value in our training set (Figure 1) and the implied scale factor to reconcile these was calculated. The scale factor for C–H bonds was then just taken to be the average of the individual scale factors. The factors for the C–C bond and the torsion angles were determined in a similar fashion. The bending frequencies are strongly coupled to other terms and thus the scale factors for angles, together with cross terms, were then determined by minimizing the root-mean-square error in all calculated vs observed frequencies using the previously determined scale factors for C–C and C–H. To determine the C–C bond reference values, the C–C bonds in the training set were compared to the Hartree–Fock values and the reference value was incremented by a value equal to the negative of the average error. The C–H bond reference was determined similarly.

Training Set Molecules and Number of Experimental Observables. The molecules used in the "training set" to derive the "experimental" energy surface from QMCFF93 are given in Figure 1. A small subset of available molecules was chosen leaving a large set to test the resulting force field since only a few parameters (7) need to be determined as described in more detail below. These molecules were selected to incorporate a wide variety of molecular environments and strain for the alkane functional

(38) Wilson, E. B., Jr.; Decius, J. C.; Cross, P. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Dover Publications: New York, 1955.

(39) Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **1991**, *12*, 186–199.

Derivation of Class II Force Fields

J. Am. Chem. Soc., Vol. 116, No. 6, 1994 2519

Table 2. Number of Experimental Observables in the Training Set

	no. of observables	
compd	bond lengths	freq^a (nondegenerate)
methane	1	9 (4)
ethane	2	18 (12)
n-butane	2	36 (36)
isobutane	2	36 (24)
cyclopropane	2	21 (14)
cyclobutane	2	30 (22)
cyclohexane	2	41 (25)
total	13	199 (137)

^{a} Some of the $3N - 6$ vibrational frequencies have not been observed experimentally and therefore were not available as "observables". The number of nondegenerate frequencies is given in parentheses.

group. They include straight chains (ethane, *n*-butane), branched chains (isobutane), and even small highly strained rings (cyclopropane, cyclobutane) as well as the relatively strain free six-membered ring (cyclohexane). The small rings, cyclopropane and cyclobutane, present a major challenge in that no energy surface has previously been derived which can account for these highly strained molecules while at the same time accounting for the higher homologs.[9] As Allinger et al. have pointed out,[9] the use of different force fields for 3-, 4-, and 5-membered rings creates problems in deriving force fields for heterocycles in that a different force field needs to be derived for each new substitution in these rings. It also creates problems in application as one must be careful in selecting force constants in new fused ring systems. As shown below, with the information we obtain from the quantum mechanical energy surface, it appears that we can determine the large anharmonicity and coupling regimes of the energy surface to account for these diverse systems with the single, Class II analytical representation of the force field.

Table 2 lists the number of experimental observables from the hydrocarbons comprising the training set (Figure 1) used to derive the seven optimal scale constants and reference values described above. Overall, the training set contains 13 independent bond lengths and 137 vibrational frequencies (corrected for removal of degeneracy) for a total of 150 experimental observables. There are many more experimental observables available but, as noted above, we save those to test the reasonableness and transferability of the force field. Previous force fields without benefit of extensive quantum data have required the use of a much larger set of the available data since they must determine all of the force constants directly. With the information provided by the quantum energy surface, instead of determining all 66 force constants in eq 1 (not including nonbond parameters) we need determine only the five scale factors and two reference values in eq 2. This is an order of magnitude reduction (66/7) in the number of parameters fit to experimental data. Thus 150 experimental observables gives a quite comfortable 20:1 ratio of observables to parameters. As a result, we have the luxury of having a large set of experimental data *outside* the training set with which to test the resulting force field.

Parameter-Observable Ratios and Class II Force Fields. Simpler hydrocarbon force fields with fewer potential constants and terms, such as the Lifson–Stern force field, required 24 parameters to be optimized[6] (small cyclic alkanes were not treated), while the MM3 force field (which includes some cross terms) contains about 50 parameters.[9] The number of parameters required to give results of comparable quality for other families of compounds increases drastically. For example, the MM3 *amide* force field employs about 200 additional parameters[39] (apart from the transferred alkane parameters) with only about 250 experimental observables available. The observable-to-parameter ratio is thus less than 2:1. By contrast, with the method used here the number of parameters remains on the order of 10 and this ratio will still be very large. The number of adjustable parameters remains roughly unchanged since only a few additional reference values and scale factors need to be optimized. By using the

Table 3. Optimized Scale Constants and Reference Values

(A) Scale Constants	
bond stretching	
S_b(C–C)	0.88
S_b(C–H)	0.83
angle bending S_θ	0.81
torsion S_ϕ	0.84
cross term S_c	0.87
(B) Reference Values for C–C and C–H Bonds (Å)	
$b^0{}_{C-C}$	1.535
$b^0{}_{C-H}$	1.111

information contained in the analytical representation of the ab initio energy, we have therefore achieved a drastic reduction in the number of parameters to be optimized against the experimental data. The corollary to this is that there is insufficient experimental data with which to derive a Class II energy surface solely from experimental observables for many heteroatomic functional groups. This is one of the major reasons for the development of this protocol in which the quantum mechanical energy surface is invoked to "amplify" the information in the experimental data.

III. Results

Scale Factors and Force Constants for Class II Alkane Force Field. The five optimized scale factors and two adjusted reference values which result from the fit to the data as described above are given in Table 3. As can be seen from Table 3, the scale factors range from 0.81 to 0.88, which is roughly in accord with what one might expect from those obtained by direct scaling of quantum mechanical frequencies.[19-29] The resulting complete set of force constants, bond and angle reference values, and nonbond parameters are given in Table 4 for the Class II force field for hydrocarbons (eq 1).

IV. Fits to Structures, Energies, and Vibrational Frequencies of Molecules in the Training Set

Molecular Structures. The results of the fit of the CFF93 force field to the equilibrium structures of the molecules comprising the training set are summarized in Tables 5 and 6. In Table 5, we present structures of acyclic compounds, while in Table 6 the results for cyclohexane and the strained rings are listed. Since only the reference values for the bond lengths were determined from the experimental structural information, the fit to the angles is a first test of the reasonableness of the force field. The quantum mechanical results are also given for comparison, as well as providing theoretical results for comparison of internals which have not been experimentally observed. All the ab initio calculations are HF/6-31G* full-geometry optimization results obtained with Gaussian 88.[40] In order to help put these results in perspective the results from two Class I force fields are also included, AMBER[12] and CVFF.[14]

(A) Acyclic Hydrocarbons. One of the first observations we can make from the results of the acyclic hydrocarbons presented in Table 5 is related to the quality of the Hartree–Fock 6-31G* basis set. As seen from this table, in general, structural variations are reproduced extremely well by the Hartree–Fock procedure, and overall, this procedure gives structures in quite good agreement with experiment. Bonds are systematically too short especially the C–H bonds which are short by about 0.02 to 0.03 Å. The angles, on the other hand, faithfully follow the experimental data. Not surprisingly, the CFF93 Class II force field does on the average an even better job of reproducing the experimental data as the systematic deviations in bond lengths have been corrected. As with the quantum mechanics, the agreement between the calculated and observed structures of these simple acyclic

(40) Frisch, M. J.; Head-Gordon, M.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; DeFrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. *Gaussian 88*; Gaussian, Inc.: Pittsburgh, PA 15213.

Table 4. Force Constants and Reference Values for the CFF93 Force Field for Hydrocarbons

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	1.111	346.4	−706.6	863.2
C–C	1.535	299.4	−515.8	667.1

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 (deg)	${}^2K_\theta$ (kcal mol^{-1} rad^{-2})	${}^3K_\theta$ (kcal mol^{-1} rad^{-3})	${}^4K_\theta$ (kcal mol^{-1} rad^{-4})
H–C–H	107.7	41.7	−7.2	−8.8
H–C–C	110.8	42.7	−8.8	−9.1
C–C–C	112.9	42.3	−9.8	−9.2

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	−0.968	0.010	−0.150
H–C–C–C	−0.968	0.010	−0.150
C–C–C–C	−0.968	0.010	−0.150

D. van der Waals Interaction

$$E = \epsilon[2(r^*/r)^9 - 3(r^*/r)^6]; \text{ Where } r^* = [(r_i^{*6} + r_j^{*6})/2]^{1/6}; \ \epsilon = (\epsilon_i\epsilon_j)^{1/2}2(r_i^*r_j^*)^3/(r_i^{*6} + r_j^{*6})$$

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

E. Bond Increment

$$E = 332.0q_iq_j/r_{ij}; \text{ Where } q_i = \sum_k \delta_{ik}$$

bond	δ_{ik} (e)
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	10.24
H–C/C–C	9.84
C–C/C–C	8.75

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})	bond/angle	$K_{b\theta}$ (kcal mol^{-1} Å^{-1} rad^{-1})
H–C/H–C–H	19.7	C–C/H–C–C	30.0
H–C/H–C–C	12.1	C–C/C–C–C	16.0

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})	angle/angle	common bond	$K_{\theta\theta'}$ (kcal mol^{-1} rad^{-2})
H–C–H/H–C–H	H–C	0.81	H–C–C/H–C–C	C–C	−1.90
H–C–H/H–C–C	H–C	1.36	H–C–C/C–C–C	C–C	−4.02
H–C–C/H–C–C	H–C	2.71	C–C–C/C–C–C	C–C	−7.81

I. Angle/Angle/Torsion

$$E = K_{\theta\theta'\phi}(\theta - \theta_0)(\theta' - \theta'_0)\cos\phi$$

angle/angle/torsion	$K_{\theta\theta'\phi}$ (kcal mol^{-1} rad^{-2})
H–C–C/H–C–C/H–C–C–H	−12.9
H–C–C/C–C–C/H–C–C–C	−16.7
C–C–C/C–C–C/C–C–C–C	−31.0

J. Bond/Torsion (Central Bond)

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

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	−45.4	−0.68	−0.78
C–C/H–C–C–C	−42.7	−2.76	−0.34
C–C/C–C–C–C	−41.1	−4.91	0.27

K. Bond/Torsion (Terminal Bond)

$$E = (b' - b'_0)[{}^1K_{\phi b'}\cos\phi + {}^2K_{\phi b'}\cos 2\phi + {}^3K_{\phi b'}\cos 3\phi]$$

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	0.68	0.41	0.07
H–C/H–C–C–C	0.68	0.25	0.35
C–C/H–C–C–C	1.45	0.51	−0.30
C–C/C–C–C–C	2.00	0.67	0.31

Table 4 (Continued)

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	−1.41	0.64	−0.39
H–C–C/H–C–C–C	0.23	0.61	−0.14
C–C–C/H–C–C–C	−1.64	0.09	−0.26
C–C–C/C–C–C–C	−0.03	−0.01	0.00

Table 5. Comparison of Experimental and Calculated Structures (Bonds in Å, Angles in deg) for Methane, Ethane, *n*-Butane, and Isobutane

compd	property	expt	ab initio	Class II CFF93[a]	Class I AMBER	Class I CVFF
methane	b(CH)	1.107(1)[b]	1.084	1.108(0.001)	1.090	1.105
ethane	b(CC)	1.534(1)[c]	1.527	1.526(−0.008)	1.530	1.521
	b(CH)	1.112(1)	1.086	1.112(0.000)	1.091	1.106
	θ(HCH)		107.7	108.0	109.1	107.0
	θ(CCH)	111.0(2)	111.2	111.0(0)	109.8	111.8
n-butane (*trans*)	b(C$_1$C$_2$)	1.531(2)[d]	1.528	1.534(0.003)	1.533	1.528
	b(C$_2$C$_3$)		1.530	1.538	1.535	1.533
	b(C$_1$H$_a$)	1.117(5) av	1.085	1.112(−0.005)	1.090	1.105
	b(C$_1$H$_b$)	1.117(5) av	1.085	1.112	1.090	1.106
	b(C$_2$H)		1.088	1.115	1.092	1.108
	θ(CCC)	113.8(4)	113.0	113.2 (−0.6)	111.0	111.6
	θ(C$_2$C$_1$H$_a$)	110.0(5) av	111.5	111.6 (1.6)	109.9	111.3
	θ(C$_2$C$_1$H$_b$)	110.0(5) av	111.0	110.8 (0.8)	109.9	112.1
	θ(HC$_2$H)		106.2	106.7	108.7	105.7
isobutane	b(CC)	1.535(1)[e]	1.531	1.537(0.002)	1.533	1.530
	b(C$_1$H)	1.122(6)	1.089	1.116(−0.006)	1.094	1.111
	b(CH)m	1.113(2)	1.086	1.113(0.000)	1.091	1.105
	θ(CCC)	110.8(2)	111.0	110.7(−0.1)	109.9	110.2
	θ(CC$_1$H)	108.1(2)	107.9	108.2(0.1)	109.0	108.7
	θ(C$_1$CH$_1$)	111.4(4)	111.3	111.6(0.2)	110.0	111.8
	θ(C$_1$CH$_2$)	110.1(3)	110.9	110.6(0.5)	110.3	112.4
	θ(H$_1$CH$_3$)	108.7(11)	107.8	107.4(−1.3)	108.7	106.6
	θ(H$_1$CH$_2$)	106.5(17)	107.7	107.7(1.2)	108.9	107.0

[a] Values in parentheses are deviations from the experiment. [b] Bartell, L. S.; Kuchitsu, K.; deNeui, R. *J. Chem. Phys.* **1961**, *35*, 1211. [c] Bartell, L. S.; Higginbotham, H. K. *J. Chem. Phys.* **1965**, *42*, 851. [d] Bradford, W. F.; Fitzwater, S.; Bartell, L. S. *J. Mol. Struct.* **1977**, *38*, 185. [e] Hilderbrandt, R. L.; Weiser, J. D. *J. Mol. Struct.* **1973**, *15*, 27.

Table 6. Comparison of Experimental and Calculated Structures (Bonds in Å, Angles in deg) for Cyclopropane, Cyclobutane, and Cyclohexane

compd	property	expt	ab initio	Class II CFF93[a]	Class I AMBER	Class I CVFF[e]
cyclopropane	b(CC)	1.514(4)[b]	1.497	1.503(−0.011)	1.527	
	b(CH)	1.099(5)	1.076	1.104(0.005)	1.090	
	θ(CCH)	117.9	118.1	118.7(0.8)	115.8	
	θ(HCH)	114.5(9)	114.0	112.7(−1.8)	119.7	
cyclobutane	b(CC)	1.552(1)[c]	1.545	1.549(−0.003)	1.528	1.589
	b(CH$_e$)	1.093(3) av	1.084	1.107 (0.016)	1.091	1.104
	b(CH$_a$)	1.093(3) av	1.085	1.111	1.091	1.106
	θ(CCC)		88.5	88.4	90.0	89.8
	θ(CCH$_e$)		117.7	118.0	112.8	113.7
	θ(CCH$_a$)		111.7	111.6	112.8	112.5
	θ(HCH)	106.4(13)	108.5	108.3(1.9)	113.5	112.7
	θ(CCCC)		18.2	19.3	0	6.2
cyclohexane (*chair*)	b(CC)	1.536(2)[d]	1.532	1.543(0.007)	1.531	1.547
	b(CH$_e$)	1.121(4) av	1.087	1.114(−0.006)	1.091	1.107
	b(CH$_a$)	1.121(4) av	1.089	1.115	1.091	1.108
	θ(CCC)	111.4(2)	111.5	111.0(−0.4)	110.2	111.3
	θ(CCH$_e$)		110.1	110.5	109.5	109.9
	θ(CCH$_a$)		109.2	109.2	109.4	109.8
	θ(HCH)	107.5(15)	106.6	106.4(−1.1)	108.8	105.9
	ϕ(CCCC)	54.9(4)	54.8	56.2(1.3)	54.8	55.1

[a] Values in parentheses are deviations from the experiment. [b] Yamamoto, S.; Nakata, M.; Fukuyama, T.; Kuchitsu, K. *J. Phys. Chem.* **1985**, *89*, 3298. [c] Egawa, T.; Fukuyama, T.; Yamamoto, S.; Tabayashi, F.; Kambara, H.; Ueda, T.; Kuchitsu, K. *J. Chem. Phys.* **1987**, *86*, 6018. These are r_z values; the r_g values as reported by Allinger et al.[9] (obtained through personal communication with Kuchitsu) give 1.554(1) Å for the CC bond, 1.109 Å for the averaged CH bonds, and 35° for the puckering angle (see the definition shown in Table 10). [d] Bastiansen, O.; Fernholt, L.; Seip, H. M.; Kambara, H.; Kuchitsu, K. *J. Mol. Struct.* **1973**, *18*, 163. [e] The coupling or cross terms in CVFF are not valid in the range of distortions contained in cyclopropane and do not yield a stable minimum for the 3-member ring. (The calculations in Table 10 were carried out without cross terms.)

molecules is extremely well reproduced by the Class II force field. These are relatively simple molecules and one might not expect to see any major differences between the Class II and diagonal quadratic force fields. The latter should reproduce molecular structures fairly well where there is little strain and anharmonicity. However, even here we can begin to see the problems that the Class I force fields have in reproducing trends. Thus, the C–C–C angle in *n*-butane and isobutane is predicted to be essentially the same by the Class I force field (AMBER), ~111°, whereas in *n*-butane the experimental value is about

Table 7. Comparison of Experimental and Calculated Conformational Energy Differences and Rotational Barriers (kcal/mol) for Simple Acyclic Hydrocarbons

	expt	MP3 (ab initio)	CFF93	MM3	AMBER	CVFF
Barriers						
ethane	2.88[a]		2.75	2.41	3.01	3.07
propane	3.40[b]		3.11		3.27	3.59
isobutane	3.9[c]		3.51		3.48	4.13
neopentane	4.2–4.8[d]		3.95	3.35	3.94	5.07
butane syn-rotation	4.56[e]		4.95	4.83	5.31	8.20
butane trans-gauche barrier	3.30[f]		3.47	3.30	3.53	4.24
Conformational Energy Differences						
butane						
trans→gauche	0.497(0.22) – 0.89(0.03)[g]	0.75	0.72	0.81	0.79	1.66
pentane						
anti-anti → anti-gauche	0.465(0.30),[h] 0.560 (0.10)[i]	0.76	0.69	0.86	0.76	1.57
anti-anti → gauche-gauche		1.36	1.38	1.62	1.48	3.10

[a] Hirota, E.; Endo, Y.; Saito, S.; Duncan, J. L. *J. Mol. Spectrosc.* **1981,** *84,* 3587. [b] Pitzer, K. S. *J. Chem. Phys.* **1944,** *12,* 310. [c] Lide, D. R., Jr.; Mann, D. E. *J. Chem. Phys.* **1958,** *29,* 914. [d] Durig, J. R.; Craven, S. M.; Harris, W. C. *Vibrational Spectra and Structure*; Marcel Dekker: New York, 1978; Vol. I. [e] Compton, D. A.; Montero, S. M.; Murphy, W. F. *J. Phys. Chem.* **1980,** *84,* 3587. [f] Pitzer, K. S. *J. Chem. Phys.* **1944,** *12,* 310. [g] Durig, J. R.; Compton, D. A. C. *J. Phys. Chem.* **1979,** *83,* 265 and references therein. [h] Maissara, M.; Cornut, J. C.; Devaure, J.; Lascombe, J. J. *Spectrosc. Int. J.* **1983,** *104,* 2. [i] Kanesaka, I.; Snyder, R. G.; Strauss, H. L. *J. Chem. Phys.* **1986,** *84,* 395.

114°, significantly larger than the 111° in the isobutane. The CVFF force field does somewhat better in this case as it does contain some of the coupling terms present in the Class II functions. Likewise, the H–C_2–H angle in normal butane is predicted to be 106° by the ab initio calculation (corresponding to the larger C–C–C angles), significantly smaller than the standard tetrahedral value. Once again, Class I force fields are unable to handle these trends resulting from intramolecular environments which reflect coupling between different coordinates, and the angle is predicted to be too large at 109°. This is essentially the same as the H–C–H angle in ethane as calculated by the diagonal quadratic force field (Amber in this case). Thus, even in these simple acyclic molecules we begin to see the problem inherent in the current generation Class I force fields which may not be able to predict what could be important variations of the target molecules that we desire to treat with these force fields. This is especially the case since the interest in novel molecules often concerns the behavior of substructural elements in a variety of intramolecular and intermolecular environments introducing different couplings and strain.

Cyclic Alkanes: Are Cyclopropane and Cyclobutane Relevant? In Table 6 we present the results for cyclohexane, cyclopropane, and cyclobutane. We would not expect the Class I force fields to be able to reproduce the properties of cyclopropane and cyclobutane since these are extremely strained molecules, and it is expected that the simple diagonal quadratic approximation will break down in this region of the energy surface. Nevertheless, the nature of the deviations is once again indicative of the inherent weaknesses in simple diagonal quadratic force fields. In addition, it should be kept in mind that *the objective is to be able to apply a force field to a wide range of new molecules.* In the attempt to design molecules with desired properties or functions, new structural moieties, new intramolecular environments, and new conformational constraints are often introduced. This is essentially the basis of molecular design. Thus, it is desirable to have a force field which is predictive and can handle a wide range of distortions. If the force field can handle distortions even beyond those which would be expected in molecular design procedures, then we will be "interpolating" in our application to the unknown target compound rather than "extrapolating" the use of the force field—the former always a safer procedure. That is, if the molecule of interest is well within the range of demonstrated validity of the energy surface in terms of internal coordinates and strain, then in general, there can be greater confidence in the force field to faithfully reflect the intramolecular forces. It is for this reason that it is important to consider molecules such as cyclopropane and cyclobutane to assess the range of validity of the energy surface and its robustness. The Class II force field and quantum mechanics both accurately reproduce the trends

resulting from the intramolecular strain in these molecules as seen for example in the C–C bond lengths. The C–C bond length in cyclopropane is only about 1.51 Å, 0.04 Å smaller than that found in cyclobutane which in turn is approximately 0.01 Å larger than that found in cyclohexane. Again, the Class I force fields are unable to reproduce the trends and tend to yield for the most part a very narrow range of angles and bond lengths. For example, the bond lengths calculated with the quadratic diagonal Class I force field for cyclopropane, cyclobutane, and cyclohexane are all within 0.003 Å of 1.53 Å, close to the reference values imposed by the harmonic forms. We will return to a more in-depth analysis of these intriguing small rings below.

Conformational Energies and Rotational Barriers. We now consider the ability of the various energy surfaces to account for the conformational energies and rotational barriers of the relatively simple molecules comprising the training set. In Table 7 the rotational barriers of ethane, propane, butane, isobutane, and neopentane are given. One sees from the first four barriers that as the chain increases in length and steric hindrance the barrier to rotation of the methyl group increases monotonically from 2.9 to roughly 4.2-4.8 on going from ethane to neopentane. The Class II CFF93 force field does a good job of reproducing both the trend and values of these barriers as, in fact, do the other force fields. MM3 does slightly less well with the barrier to neopentane (3.35 vs 4.2–4.8) but overall there is no clear distinction between the different energy surfaces. Part of the reason for this is the simplicity of these molecules and the fact that the Class I force fields were parametrized in many cases against these energy barriers. When molecules outside the training set of the Class I molecules and with more complicated intramolecular environments are addressed, the inability of the Class I functional form to respond to environmental effects is also reflected in the energetics (to be published).

Discrepancy between theory and experiment for the *syn* rotational barrier of *n*-butane has generated discussion over the years. The spectroscopically determined value of Compton and co-workers[41] for this barrier is 4.56 kcal/mol (quoted as the experimental data in the present work), while earlier ab initio calculations[42] give values of about 6 kcal/mol for this barrier. However, recently Allinger, Schaefer, and colleagues[43] showed that using a larger basis set with electron correlation decreases the ab initio value by about 1 kcal/mol, thus bringing theory and experiment into better agreement. Our calculated value for this barrier is 4.95 kcal/mol, in good agreement with both experiment

(41) Compton, D. A.; Montero, S. M.; Murphy, W. F. *J. Phys. Chem.* **1980,** *84,* 3587.

(42) Raghavachari, K. *J. Chem. Phys.* **1984,** *84,* 1383.

(43) Allinger, N. L.; Grev, R. S.; Yates, B. F.; Schaefer, H. F., III *J. Am. Chem. Soc.* **1990,** *112,* 114.

Derivation of Class II Force Fields

J. Am. Chem. Soc., Vol. 116, No. 6, 1994 2523

Table 8. Root-Mean-Square (rms) and Maximum (max) Deviations (cm⁻¹) from Experiment of CFF93 Calculated Vibrational Frequencies Used in the Training Set and Comparison with Results from AMBER, CVFF, and MM3

| | rms deviations (cm⁻¹) | | | | maximum deviations (cm⁻¹) | | | |
| | Class II | | Class I | | Class II | | Class I | |
compd	CFF93	MM3	AMBER	CVFF	CFF93	MM3	AMBER	CVFF
methane	28		73	38	37		-134	-94
ethane	22	58	93	30	-38	-126	-231	-81
n-butane	25	41	78	31	61	-128	-207	76
isobutane	18	30	70	19	38	-85	-162	46
cyclohexane	23	32	76	34	59	-117	-185	86
cyclobutane	41		102	81	-97		-234	-193
cyclopropane[a]	51	132	173	174	126	365	-335	342
	(77)	(155)	(197)	(208)	(-204)	(370)	(-443)	(-431)
average[a]	30 (34)	59 (63)	95 (98)	58 (63)				

[a] Results in parentheses include two CH_2 modes (rock and wag) in cyclopropane (see text).

Table 9. Comparison of Experimental and Calculated Vibrational Frequencies for Cyclopropane

| | | Class I force fields | | Class II force fields | |
mode	expt[a]	AMBER	CVFF[b]	CFF93	MM3
CH str (A₂″)	3102	2955 (-147)	2983 (-119)	3052 (-50)	3082 (-20)
CH str (E″)	3083	2965 (-118)	2995 (-88)	3049 (-34)	3079 (-4)
CH str (E″)	3083	2965 (-118)	2995 (-88)	3049 (-34)	3079 (-4)
CH str (A₁′)	3038	2865 (-173)	2900 (-138)	3009 (-29)	3015 (-23)
CH str (E′)	3024	2848 (-176)	2879 (-145)	3001 (-23)	2990 (-34)
CH str (E′)	3024	2848 (-176)	2879 (-145)	3001 (-23)	2900 (-34)
CH₂ bend (A₁′)	1482	1736 (254)	1824 (342)	1451 (-31)	1519 (37)
CH₂ bend (E′)	1438	1618 (180)	1712 (274)	1456 (18)	1370 (-68)
CH₂ bend (E′)	1438	1618 (180)	1712 (274)	1456 (18)	1370 (-68)
CH₂ rock (E″)	1187	1174 (-13)	1303 (116)	1232 (45)	1426 (239)
CH₂ rock (E″)	1187	1174 (-13)	1303 (116)	1232 (45)	1426 (239)
ring breathing (A₁′)	1188	1174 (-14)	1203 (15)	1314 (126)	1209 (21)
CH₂ twist (A₁″)	1126	1007 (-119)	1121 (-5)	1017 (-109)	761 (-365)
CH₂ wag (E′)	1028	1003 (-25)	1028 (0)	1048 (20)	1097 (69)
CH₂ wag (E′)	1028	1003 (-25)	1028 (0)	1048 (20)	1097 (69)
CH₂ wag (A₂′)	1070	627 (-443)	685 (-385)	875 (-196)	930 (-140)
ring defm (E′)	869	696 (-173)	766 (-103)	831 (-38)	803 (-66)
ring defm (E′)	869	696 (-173)	766 (-103)	831 (-38)	803 (-66)
CH₂ twist (E″)	738	403 (-335)	445 (-293)	711 (-27)	833 (95)
CH₂ twist (E″)	738	403 (-335)	445 (-293)	711 (-27)	833 (95)
CH₂ rck (A₂″)	854	394 (-210)	423 (-431)	650 (-204)	1224 (370)
rms		173 (197)[c]	174 (208)	51 (77)	132 (155)
max		-335 (-443)	342 (-431)	126 (-204)	-365 (370)

[a] Duncan, J. L.; Burns, G. R. *J. Mol. Spectra.* **1969**, *30*, 253. [b] The calculation with CVFF does not include cross terms as noted in Table 6, footnote *e*. CVFF is classified here as a Class I force field for its inadequacy for small rings. [c] Results in parentheses include the two anomalous CH_2 frequencies.

and the recent theory, as are the values calculated by MM3 and AMBER. CVFF produces a value some 3 kcal too high for this barrier and is also high for some of the conformational energy differences as we shall see below. The smaller barrier (*trans* to *gauche*) in *n*-butane is also accurately predicted by CFF93 as well as the other three force fields.

Conformational Energies. There are also three conformational energies available in the molecules within the training set involving trans-gauche conformations. As seen in Table 7 these are also well accounted for by most of the force fields. In butane the experimental trans-gauche energy difference ranges from 0.5 to 0.9 kcal[44] and it is given by quantum mechanics as 0.75. Most of the potential functions yield results on the high side of the experimental range in agreement with the ab initio results. The same is true of *n*-pentane where the anti-anti to anti-gauche energy experimentally ranges from 0.46[45] to 0.56,[46] and again quantum mechanics gives a value of about 0.76.[47] The force fields are once again in agreement with the larger value. No experimental values are available for the anti-anti to gauche-gauche conformational energy difference, but quantum mechanics gives a value

of about 1.36 which is in reasonable agreement with MM3 and CFF93. CVFF as with the *syn* rotational barrier gives somewhat high values for all three of the conformational energies.

Vibrational Frequencies. In Table 8 we turn to the ability of the various energy surfaces to reproduce vibrational frequencies. This property is important for several reasons. Foremost, the vibrational frequencies are dynamic properties and reflect the ability of the force field to account for the dynamics of molecules. Secondly, the vibrational frequencies reflect the curvature of the energy surface itself. Thus, to the extent that the vibrational frequencies are not accounted for correctly, they indicate that the curvature of the potential surfaces is in error, which in turn implies that at some point further along, the value of the energy surface itself will be in error (i.e., the potential energy and forces are in error). As we can see from Table 8, which gives the rms and maximum deviations of the calculated frequencies for each of the molecules included in the training set, the Class II CFF93 functional form reproduces the frequencies of these molecules extremely well with the exception of the CH_2 rocking and wagging modes in cyclopropane. The latter modes are characteristically anomalous in all force fields with deviations of 200–400 cm⁻¹ in the different energy surfaces including the MM3 cyclopropane force field which was constructed solely for cyclopropanes[48] (Table 9). Clearly these modes warrant further study. Excluding these modes, the overall rms deviation in CFF93 is 30 cm⁻¹.

(44) Durig, J. R.; Compton, D. A. C. *J. Phys. Chem.* **1979**, *83*, 265 and references therein.

(45) Maissara, M.; Cornut, J. C.; Devaure, J.; Lascombe, J. *J. Spectrosc. Int. J.* **1983**, *104*, 2.

(46) Kanesaka, I.; Snyder, R. G.; Strauss, H. L. *J. Chem. Phys.* **1986**, *84*, 395.

(47) Wiberg, K. B.; Murko, M. A. *J. Am. Chem. Soc.* **1988**, *110*, 8029.

Table 10. Comparison of Geometries, Vibrational Frequencies (ν), and Ring Puckering (ϕ) of Cyclobutane and Cyclopropane As Computed with CFF93, AMBER, CVFF, and MM3

	expt[b]	CFF93	AMBER	CVFF[d]	MM3
			Cyclobutane		
b(CC) (Å)	1.552	1.549(−0.003)	1.530(−0.024)	1.531(−0.021)	1.557(0.005)
ϕ (deg)[a]	27.9	27.5(−0.4)	0(−27.9)	0(−27.9)	32.3(4.4)
θ(CCH$_e$) (deg)	(117.7)[c]	118.0	112.8	113.7	
θ(CCH$_a$) (deg)	(111.7)[c]	111.6	112.8	112.5	
ν_{rms} (cm^{-1})		41	102	81	
ν_{max} (cm^{-1})		97	234	193	
			Cyclopropane		
b(CC) (Å)	1.512	1.503(−0.009)	1.527(0.015)	1.529(0.017)	1.512(0.000)
θ(CCH) (deg)	(118.1)[c]	118.7	115.8		117.4
ν_{rms} (cm^{-1})		77	213	205	155
ν_{max} (cm^{-1})		205	460	431	370

[a] Puckering angle. [b] See Table 6 for experimental references. [c] Quantum mechanical result. [d] The cross terms were omitted from the calculation here as cyclopropane is predicted to be unstable with these cross terms (see footnote in Table 6). Results for cyclobutane calculated with the cross terms included are given in Table 6.
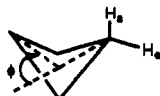
For the most part the rms deviations in frequency are in the range of 20–30 cm^{-1} with cyclobutane and cyclopropane being slightly higher at ~40–50 cm^{-1} (excluding the problematic modes in the latter). We note that for a subset of these molecules including ethane, methane, *n*-butane, isobutane, and cyclohexane the best fit to frequencies achieved to date in the literature is due to the force field of Lifson and Stern.[6] The authors also found that the keys to the improvement of frequency and improvement of the force field included the introduction of anharmonicity in the form of a Morse potential[5] as well as correlated cross terms. It is interesting to speculate on why these authors were able to achieve a better fit for these molecules. In fact, this may be the first indication of the limitation of the scaling technology. Thus, by being able to fit separately the two stretching potentials, as well as the two parameters in the Morse potential, and adjusting the C–C–C, C–C–H, and H–C–H parameters separately rather than insisting that they be related as they are in the quantum mechanics, one may achieve an improved fit to the extent of some 12 cm^{-1}. However, another possibility is that since these authors did not calculate energies, energy barriers, and structures it may be that a better fit to the frequencies was achieved at the expense of these other properties. We see from Table 8 that the rms deviation in frequencies for the two Class I force fields are ~100 and 60 cm^{-1}, once again reflecting the limitations inherent in this functional form for accounting for the shape of molecular energy surfaces. In addition to the AMBER and CVFF results, Allinger has mentioned that calculations with MM2 would have a rms deviation of about 80 wavenumbers for *normal* alkanes.[10] The deviation would, of course, be much larger if one tried to apply MM2 to the small cyclic compounds. Thus, again we see the limitations of the Class I functional forms, where even though fit somewhat less well, the energy surfaces of cyclopropane and cyclobutane are still accounted for as well as or better by the Class II energy surface than even unstrained hydrocarbons are by the quadratic diagonal representation.

Small Cyclic Rings. As noted above, a hallmark of the CFF93 force field is the relatively small number of atom types, which greatly extend its transferability and hence the range of species to which it may be applied accurately. Its ability to fit the small ring molecules such as cyclobutane and cyclopropane without introducing special parameters provides a severe test of this transferability. Here we expand on our analysis of the results on these molecules with both the conventional diagonal force fields (AMBER[12] and CVFF[14]) as well as with CFF93. Some relevant results are extracted from the previous discussion and summarized in Table 10, which gives the calculated C–C bond distances, equatorial and axial C–C–H angles, and puckering angle of cyclobutane and the rms and maximum deviations of vibrational

frequencies. As can be seen from this table, the Class II force field fits these strained cyclic molecules extremely well with only a slight degradation in reproduction of vibrational frequencies. In particular, it accounts for the reduction in the C–C bond length of cyclopropane from approximately 1.53 Å in the acyclic molecules to 1.50 Å in cyclopropane, while at the same time accounting for the increase in the bond length in cyclobutane. It also accounts reasonably well for the large distortion of the C–C–H angle, which opens to 118° in cyclopropane according to the ab initio results. In cyclobutane, there is a 6° opening of the equatorial C–C–H angle relative to the axial CCH value (117.7° vs 111.7°) which is again reproduced well by the Class II force field. While the rms errors in the calculated vibrational frequencies with CFF93 are 41 and 77 cm^{-1} for cyclobutane and cyclopropane, respectively (the larger rms deviation for cyclopropane is largely due to the 205-cm^{-1} deviation in a concerted methylene vibration mode as described in ref 48), with AMBER and CVFF the corresponding rms errors are 102 and 81 cm^{-1} for cyclobutane and 197 and 208 cm^{-1} for cyclopropane. Thus the errors in the CFF93 frequencies are consistently about 45% less than CVFF and 60% less than AMBER. As an additional example, even with parameters derived specifically for cyclopropane, MM3's rms and maximum frequency deviations for cyclopropane are 155 and 370 cm^{-1}, respectively.[48] Although cyclopropane and cyclobutane frequencies are fit better with the Class II functions than even the normal alkanes by the Class I functional form, we note the slight degradation in fit to frequencies (of ~10 cm^{-1}) for the small rings relative to the remaining alkanes and the unexplained large deviation in the CH$_2$ rock. This is the first indication of what might be a subtle deficiency in the energy surface in these highly distorted regimes, i.e., a possible indication of a missing coupling or breakdown in transferability. This is discussed further below. The source of the large deviation for the methylene vibration and the slight degradation in fit are also subjects for further study.

Puckering of Cyclobutane. Perhaps the most dramatic difference in the energy surfaces is their ability to reproduce the puckered geometry of cyclobutane. As seen in Table 10, rather than being planar cyclobutane is observed to be significantly puckered with a puckering angle, ϕ, of 28° (for a definition of ϕ see Table 10 footnote). The diagonal quadratic Class I force fields cannot account for this and and predict cyclobutane to be planar. CVFF, which contains some cross terms, in fact predicts a slight pucker (see Table 6). Even in MM3 a special torsion parameter is required to achieve this pucker. Note this torsion in MM3 implies that the barrier to rotation about the single C–C bond is 20.7 kcal/mol in cyclobutane as opposed to 4.68 kcal/

(48) Aped, P.; Allinger, N. L. *J. Am. Chem. Soc.* **1992**, *114*, 1–16.

Table 11. Root-Mean-Square (rms) Deviations (in cm⁻¹) from ab Initio (HF/6-31G*) in the Calculated Vibrational Frequencies of Two QM Force Fields[a]

molecule (conformation)	QMFF A fits normal alkanes only	QMFF B fits small rings also	B − A
I. Normal Alkanes			
methane	26.9	31.4	4.5
ethane (staggered)	16.1	22.6	6.5
(eclipsed)	27.6	30.0	2.4
propane	17.1	20.0	2.9
butane (trans)	15.9	18.5	2.6
(gauche)	15.9	18.6	2.7
pentane (trans-trans)	15.2	19.0	3.8
(trans-gauche)	14.2	16.6	2.4
isobutane	18.4	25.3	6.9
isopentane	15.5	21.2	5.7
neopentane	30.1	31.6	1.5
cyclohexane	19.6	21.1	1.5
average	19.4	23.0	3.6
II. Small Rings			
cyclopentane (twist)	24.5	20.7	−3.8
(envelop)	23.8	22.7	−1.1
cyclobutane (puckered)	48.6	38.8	−9.8
(planar)	60.5	46.5	−14.0
methylcyclobutane (eq)	32.5	26.4	−6.1
(ax)	44.1	32.6	−11.5
cyclopropane	128.1	85.5	−42.6
methylcyclopropane	91.1	70.2	−20.9
1,2-dimethylcyclopropane	68.5	57.0	−11.5
average	58.0	44.5	−13.5

[a] A: the ab initio training set consists of only non-small-ring compounds. B: includes small rings also—and the differences between the two.

mol in normal alkanes. We see here that in addition to the other observables this puckering angle is also reproduced by the single, transferable Class II force field with no further modification.

Impact of Inclusion of Small Rings on Overall Energy Surface and Fit. As noted above, although the overall fit to the frequencies of the small rings, 40 and 50 cm⁻¹, is even better than has been achieved for normal hydrocarbons by many previous force fields, the deviations are still larger than the fits to the normal alkanes. In any case, the impact of inclusion of small rings on the resulting overall energy surface and fit to the remaining molecules is an important issue. The use of quantum energy surfaces allows us to do a straightforward experiment to address this issue. We have derived a second *quantum* force field by fitting the quantum mechanical energy surface of all molecules excluding the small rings by the methodology outlined here and in ref 1. Thus we can directly probe the impact the small rings have both on the force field itself and on the overall fit. The vibrational frequencies calculated with the two quantum force fields for a number of normal alkanes and small rings are compared to the ab initio results, and the respective rms deviations are calculated. These are reported in Table 11. From Table 11 one sees that the degradation in the calculated vibrational frequencies for normal alkanes resulting from the demand that a consistent transferable force field also fits small rings is rather small, only 3.6 cm⁻¹ on average. On the other hand, by including the small rings in the quantum energy surface and deriving a consistent transferable force field the fit to these rings improves dramatically, in some cases by over 40 cm⁻¹ in the case of cyclopropane. Similar minor improvements are obtained in the other properties. In addition, there are some force constants, mainly cross terms, which are much better defined.

Thus it is felt that this is a very small price to pay for a transferable force field which does not have the intrinsic problems associated with it of having to derive a separate force field for 3- and 4-membered rings for every class of compounds. We should emphasize, however, that there are still errors that exist within this force field as pointed out above and certainly improvements are still in order. Again, this follows Lifson's philosophy inherent

in the CFF methodology[6] of gradually perfecting the various parts of a force field by systematic probes of its deficiencies. This is being done continuously, and such an analysis has already led to the suggestion of an improved function form which takes better account of symmetry of the internal deformations (Li and Hagler, work in progress).

V. Summary

We have presented a methodology for deriving a Class II force field in which experimental data are leveraged by the extraction of the functional form and relative values of force constants from quantum mechanical energy surfaces. This has resulted in a dramatic decrease in the number of parameters that need to be determined from the experimental data allowing us to extract a highly anharmonic coupled force field from a limited number of experimental data. Only seven scaling factors and reference values were determined here from 150 observables. The resulting Class II force field was shown to be able to reproduce trends in the molecular geometries of small acyclic and cyclic hydrocarbon molecules caused by intramolecular environmental effects. The latter are not accounted for as well by the simpler quadratic diagonal force fields, or Class I force fields, commonly used in molecular mechanics simulations of complex organic and bio-molecular systems. Perhaps the most dramatic improvement displayed by this energy surface is its ability to account for the dynamics of the diverse set of molecules in the training set. Vibrational frequencies are fit to 20–30 cm⁻¹ for cyclohexane and acyclic hydrocarbons and even the small rings are fit to within ~40–50 cm⁻¹. Typical deviations for these frequencies calculated with a Class I energy surface are more than twice this, ~70–80 cm⁻¹ for acyclics and 100–200 cm⁻¹ for small rings. It was also shown that the single Class II force field, which includes anharmonicity and accounts for intramolecular coupling interactions, becomes much more transferable and *for the first time a single consistent force field was shown to be able to account for the properties even of such strained small rings as cyclopropane and cyclobutane* as well as larger rings and acyclic molecules.

Thus, the single Class II force field given in eq 1 derived from the quantum mechanics and scaled as in eq 2 is able to account for the properties of this diverse training set. We note again that the Hartree–Fock procedure also accounts well for the properties of these molecules including the puckering of cyclobutane, demonstrating the ability of the ab initio method to provide information on the nature of the alkane energy surface. We also note that since the force field derived from the quantum energy surface fits the quantum mechanical properties and the simple scaled force field fits the experimental trends in all these molecules, we conclude that *to the degree of fit achieved here*, there is a single energy surface from which properties of the small cyclic compounds as well as acyclic compounds arise. The simple quadratic diagonal force fields which do not account for the anharmonicity and coupling inherent in the experimental data are unable to reproduce these important trends in molecular properties induced by intramolecular environments.

Finally perhaps one of the most significant aspects of the paradigm of force field development proposed here is that it allows for the derivation of a high quality force field for any arbitrary functional group even where experimental data are sparse or nonexistent but where quantum mechanical energy first and second derivatives can be calculated. This paradigm allows a force field to be obtained if need be for a single molecule by generating the energy first and second derivatives of a sufficient number of distorted configurations. This has obvious import to the ability to treat arbitrary and newly designed molecules.