Molecular Mechanics Force-Field Parameterization Procedures

A. J. Hopfinger

Department of Medicinal Chemistry and Pharmacognosy, University of Illinois at Chicago, Chicago, Illinois 60680,

and

Department of Medicinal Chemistry, Searle Research and Development, Skokie, Illinois 60077

R. A. Pearlstein

Department of Medicinal Chemistry, Searle Research and Development, Skokie, Illinois 60077
Received 9 February 1984; accepted 16 April 1984

A set of procedures and guidelines are presented for the estimation of bond length, bond angle, and torsional potential constants for molecular mechanics force fields. The force field constants are ultimately derived by "subtracting" nonbonded molecular mechanics energies from corresponding molecular orbital energies using a model compound containing the chemical structure to be parameterized. Case study examples of bond length, bond angle, and torsional rotation force field parameterizations are presented. A general discussion of molecular mechanics force field parameterization strategy is included for reference and completeness. Finally, a curve-fitting program to generate force field parameters from raw data is given in Appendix I.

INTRODUCTION

Application of free valence geometry molecular mechanics[1] to conformational studies has often been limited due to a lack of bonding geometry and torsional force-field parameters. Most numerous among the "missing" parameters are those relating to torsional rotations. Torsional rotation parameters are also needed, and again, often not available, in fixed valence geometry (constant bond lengths and angles) molecular mechanics conformational analyses.[2] Thus there is a pressing need to devise methods and strategies to estimate force-field parameters. Workers have reported deriving parameters by matching experimental data, most notably crystal structures and energies and/or vibrational spectra, and/or microwave spectra.[3,4] Force-field parameters have also been estimated from quantum mechanical calculations, as well as a combination of experimental and computational (quantum mechanical) data.[5(a)]

The purpose of this article is to report a set of computational procedures we have devised which allows a user to add parameters to a force-field data base using selectable molecular and quantum mechanics methods. From both a practical point of view, as well as for overall consistency, these procedures should be used to expand existing parameter sets and not to initiate derivation of entirely new force fields. The consistent application of the same molecular mechanics formalism and the same molecular orbital theory in development of the force-field parameter data base should insure self-consistency and reliability in the use of the resultant force field at a level near that of the quantum mechanical method.
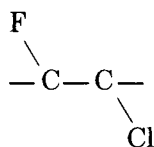
The precision of parameter estimates needs to be commensurate with their end-point application. The parameterization procedures described here are directed to estimating reasonable molecular geometries and not more detailed molecular properties such as vibrational spectra, heats of formation, and heat capacities. This should be kept in mind when employing the resultant parameters. The article has been written in something of a "cookbook" fashion so that an investigator not an expert in computational chemistry can employ the procedures in, hopefully, a straightforward fashion. This has been done in response to the increasing number of "nonexperts" using molecular mechanics software.
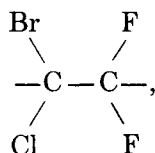
METHODS AND RESULTS

A. Generating a Model Compound

The general strategy is to first select a model compound containing the structural unit (bond length, bond angle, or torsional rotation) whose

parameters are sought. The selection of the model compound is based upon replacing all multiple valence atoms bonded to the structural unit in question with protons. Parameters for ring systems are derived from nonring model compounds in order to be in a position to estimate ring-strain effects. Further, a distinction between identical structural units that have different bonded monovalent hetero-atoms is not normally made although this can be done. The default representative model compound of each structural unit is the all protonated member. Thus, for example, the force-field parameters for a $C(sp^3)$ —$C(sp^3)$ bond are the same for

$$
\begin{array}{c}
F \\
\backslash \\
-C-C- \\
\quad\quad \backslash \\
\quad\quad\quad Cl
\end{array}
$$

as for

$$
\begin{array}{c}
Br \quad\quad F \\
\backslash \quad\quad / \\
-C-C-, \\
/ \quad\quad \backslash \\
Cl \quad\quad F
\end{array}
$$

and the default model compound is ethane. Torsional potentials about the C—C bond, on the other hand, would be unique for each of these structures since four atoms are needed to define a torsional rotation. The justifications for imposing simplifyig constraints are

(1) The valence geometry and torsional force-field parameters of a structural unit are relatively independent of the atoms bonded to the unit. The effects of the bonded atoms are almost exclusively manifested through the nonbonded atomic pair interactions.

(2) Virtually every proposed molecular mechanics force field assumes the valence geometry and torsional energetics to be independent of bonded atoms beyond the structural unit in question.

(3) The number of parameter sets would be hopelessly large unless simplifications are made.

Clearly hydrogens cannot be used in a case where a nonsingle bond is involved. In this situation, the appropriate nonhydrogen atom, the default being $C(sp^2)$, is used and it is protonated to fill out its valence requirement. Examples of this class are presented as part of Table I.

The next step in computing the parameters is to assign an initial valence geometry to the model compound. "Standard" bond lengths and angles as found, for example, in the *Handbook of*

Chemistry and Physics provide a comprehensive and reliable data base for generating valence geometry. Crystal structure geometry coupled with standard hydrogen-containing bond lengths and angles represents an alternate, or, in some cases, is a complementary approach.

In order to calculate bond length or bond angle force-field parameters, a sterically allowable conformation must first be determined. This is most readily achieved by performing a fixed valence geometry molecular mechanics calculation in which torsional rotations about single bonds are systematically explored and the corresponding conformational energetics calculated. This is normally a two-step process employing the steric, electrostatic, and hydrogen-bond potentials which will ultimately be used to complement the sought after bond length and angle force-field constants. First, conformer energy minima are sought using only the steric potential contribution. The atomic charge density distribution is then determined using one of the minimum energy conformers, normally the lowest energy state (the global energy minimum). The atomic charge densities are necessary in order to employ the electrostatic and hydrogen bond potentials. The method of calculating atomic charge density should be the same molecule orbital formalism as is to be used to determine the force-field parameters. That is, the quantum-based molecular parameters used in a molecular mechanics calculation should be derived from a common molecular orbital method. The global energy minimum conformer realized in a second conformational analysis employing electostatic and hydrogen-bonding potentials, as well as the steric term, is normally used in generating bond length and angle parameters. However, other sterically allowed conformers can be adopted if there are some secondary advantages. Torsional parameterization is not based upon a single conformation and, consequently, the strategy described here is not relevant.

The nagging concern of molecular dielectric representation encompasses parameterization strategy. For better or worse, this problem is largely resolved by the fact that the molecular orbital software packages have been written by assuming a "molecular dielectric constant" of unity. Consequently, the molecular mechanics calculations, by analogy, can be carried out using a molecular dielectric constant of unity. This working solution overlooks many complications including the representation of the dispersion of

charge density over space (molecular orbital theory) by point charges located at atomic centers (molecular mechanics model). Clearly the point charge representation increasingly breaks down as the charge distribution becomes more attentuated and/or the molecule is assigned a net nonzero charge.

B. Bond Length and Angle Parameterization
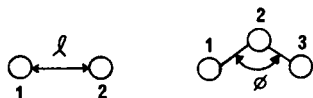
The bond length and angle force-field functions almost universally are represented by a Hookean function,

$$P(l) = k_l(l - l_0)^2 \tag{1}$$

$$P(\phi) = k_\phi(\phi - \phi_0)^2 \tag{2}$$

where P is the energy, l the bond length, ϕ the bond angle, k_l and k_ϕ the corresponding force constants, and l_0 and ϕ_0 the equilibrium bond length and angle values. The symbol μ will be used for both l and ϕ henceforce to save space.

The μ_0 can be assigned in two ways. "Standard" values of μ_0 can be selected from x-ray data and/or compiled structure lists[6-8] as discussed earlier. The alternative is to compute μ_0 by selecting a series of values of μ and computing both the molecular orbital and molecular mechanics energies of the model compound in the conformation deduced in the manner described above. $P(\mu)$ is the molecular orbital energy minus the molecular mechanics energy for each μ using the model compound. The initial set of values chosen for μ can be based upon bracketing a standard value. However, two additional criteria must be met by the final series of values of μ. First, $P(\mu)$ should be minimized as a function of μ within the set. The corresponding value of μ can be the required μ_0. Generally $P(\mu_0)$ will be subtracted from the final set of $P(\mu)$ used in the curve fitting. This is done for mathematical convenience. Second, μ_0 should be surrounded by at least two, and even better, three smaller and larger values, respectively, of μ in rounding out the set of $(\mu, P(\mu))$ employed to estimated k_μ. Generally, seven pairs of $(\mu, P(\mu))$ should be employed in calculating k_μ.

The final selection of the set of μ used to compute μ_0 and k_μ cannot be defined by any general recipe. However, the strategy is to maximize the range of μ about μ_0 so that the boundary points in this range have $P(\mu)$ values as large as possible relative to $P(\mu_0)$. However, the boundary points in μ are constrained in selection by the requirement that all corresponding $P(\mu)$ belong to a parabolic functional regime.

Overall this reduces to a trial-and-error procedure in which a plot of μ vs. $P(\mu)$ allows definition of the range in μ for which $P(\mu)$ is parabolic. In turn, this allows the set of $(\mu, P(\mu))$ for function fitting to be selected. Once the set of $(\mu, P(\mu))$ have been determined, least-squares fit procedures from any of a number of statistical packages, such as MINITAB,[9] can be used to compute (k_μ, μ_0). The direct least-square parabolic fit for eqs. (1) and (2) are

$$k_\mu = \left\{ \left[n\sum_i \mu_i^2 - \left(\sum_i \mu_i\right)^2 \right] \sum_i \mu_i^2 P(\mu_i) \right.$$

$$- \left[n\sum_i \mu_i^3 - \sum_i \mu_i^2 \sum_i \mu_i \right] \sum_i \mu_i P(\mu_i)$$

$$\left. + \left[\sum_i \mu_i^3 \sum_i \mu_i^1 - \left(\sum_i \mu_i^2\right)^2 \right] \sum_i P(\mu_i) \right\} / D \tag{3}$$

$$\mu_0 = \left\{ \left[n\sum_i \mu_i^3 - \sum_i \mu_i^2 \sum_i \mu_i \right] \sum_i \mu_i^2 P(\mu_i) \right.$$

$$- \left[n\sum_i \mu_i^4 - \left(\sum_i \mu_i^2\right)^2 \right] \sum_i \mu_i P(\mu_i)$$

$$\left. + \left[\sum_i \mu_i^4 \sum_i \mu_i - \sum_i \mu_i^4 \sum_i \mu_i^2 \right] \sum_i P(\mu_i) \right\} /$$

$$(2k_\mu D) \tag{4}$$

where

$$D = \sum_i \mu_i^4 \left[n\sum_i \mu_i^2 - \left(\sum_i \mu_i\right)^2 \right]$$

$$- \sum_i \mu_i^3 \left[n\sum_i \mu_i^3 - \sum_i \mu_i^2 \sum_i \mu_i \right]$$

$$+ \sum_i \mu_i^2 \left[\sum_i \mu_i^3 \sum_i \mu_i - \left(\sum_i \mu_i^2\right)^2 \right] \tag{5}$$

with i extending over the set of n observations of μ in which μ is either l or ϕ in eqs. (1) and (2), respectively. Appendix I lists a BASIC program that computes (k_l, l_0) and (k_ϕ, ϕ_0) from data sets.
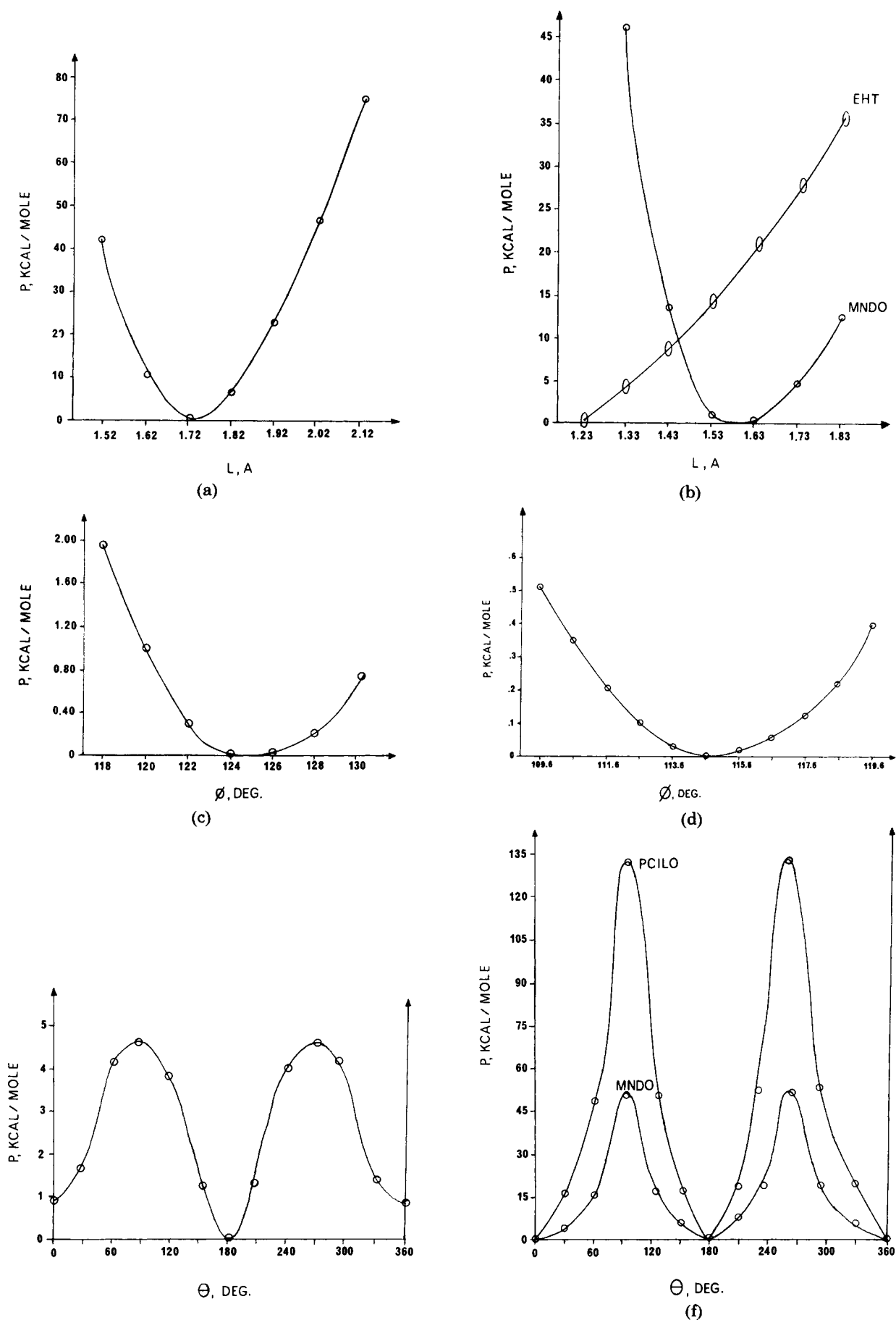
Figure 1. Energy (kcal/mol) versus distance (Å) or angle (degrees) for the case study examples of force-field parameterizations detailed in Table I. (a) Case 1, (b) case 2, (c) case 3, (d) case 4, (e) case 5, and (f) case 6.

The values of l_0 and ϕ_0 realized from the quantum mechanical analysis may not be as reliable, depending on the formalism used, as experimental values. Thus an operating procedure is to use the k_l and k_ϕ found from the quantum calculations and, as appropriate, corresponding l_0 and ϕ_0 from experimental (x-ray) sources. Some force fields require bond moments as part of the set of bond length parameters. A bond moment (BM) can be estimated by computing the charge densities for the atoms $(1, 2)$ forming the bond using l_0 as the bond length and applying eq. (6).

$$\text{BM} = 2.385 l_0 \left(Q_1 - Q_2 \right) \qquad (6)$$

where $Q_1 \geq Q_2$ and the factor 2.385 converts BM to Debye units when l_0 is in angstroms and the charges are in electron units.

Table I contains examples of bond length, bond angle, and torsional angle parameterizations. Figure 1 contains the plots of energy versus geometric variable for each parameterization study. These examples are randomly selected in that they arose in our research program over the course of writing this article.

Different quantum mechanical methods have been used in the examples to illustrate generality of the methodologies. Also, different approaches to selecting the geometries of the model compounds are reported. The specific force constants derived should not be used to augment existing force fields. These examples have been selected to provide an overview of the methodology.

Case 1 for a $C(sp^2)$—S bond is an example where the model compound cannot be generated by total protonation of the atoms forming the bond. A $=CH_2$ must be bonded to the $C(sp^2)$ atom. There is virtually free rotation about the C—S bond with respect to nonbonded interactions. The C=C bond requires the three protons and sulfur to be coplanar. Hence the planar conformation illustrated was adopted for the model compound.

Case 2 is an example of a bond length parameterization where the model compound is sufficiently simple as to not contain a torsional bond. Both MNDO[10] and EHT[11] molecular orbital approximations were used to illustrate the variability of $P(l)$ vs. l for different molecular orbital methods. Clearly, EHT is inadequate for deriving force field parameters for the S=O bond. However, experimental S=O bond lengths are 1.45–1.49 Å when the sulfur is bonded to carbons and

1.37 when bonded to fluorines.[12-14] This is a situation where the most appropriate experimental bond length should be substituted for the calculated l_0.

In case 3 bond angle parameters are required for a C(carbonyl)—$C(sp^2)$=$N(sp^2)$ group which is part of a ring system. Hence in the model compound a planar conformation has been selected which has two protons cis to mimic the geometry of two non-proton atoms in the actual ring system. This bond angle is quite variable according to experimental data (a range of $120°$–$129°$ has been reported).[8,15] The computed equilibrium value of $125.2°$ falls in the middle of this range.

Case 4 involves parameterization of one of the bond angles of an $C(sp^3)$ atom. These bond angles are quite flexible and x-ray values are reported ranging from $109°$ to $115°$.[8] However, for the specific bonbd angle under consideration, the reported upper value for ϕ is $113°$.[8] Thus the computed equilibrium value, ϕ_0, of $114.6°$ is probably too large. An experimental value of $113°$ might be a better choice for ϕ_0.

C. TORSIONAL POTENTIAL PARAMETERIZATION
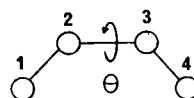
There is a very major difference in the application of torsional potential functions as compared to bond length and angle potentials. Only small excursions from the minimum in potential energy are realized in the parabolic bond length and angle potentials. Thus the emphasis in fitting these potentials is to reproduce potential-well shape near the minimum. One need not be concerned with the functional form, or parameterization, for large excursions from equilibrium because they are energetically disallowed.

Torsional potentials, on the other hand, must be evaluated over the full $360°$ of rotation about a bond. Consequently, the torsional parameters must be valid for full torsional rotation. Unless a very complex torsional function is selected (which is usually not done), the precision of fit about individual torsional minima is partially sacrificed for overall fit. The goal is to reproduce all extrema with less concern for specifying well shape about minima. In practice there can be a trade-off

between well shape and barrier height fitting. For example, in the torsional parameterization of bonds which have considerable double (or higher) bond character, where barrier heights are large, more attention should be given to well shape, at the expense of reproducing barrier heights, because major excursions from torsional minima will be energetically unlikely.

The model compounds used to derive torsional parameters are selected in the manner described earlier in Section A. However, a sterically allowed conformation of the model compound is not sought. Rather the model compound is set into a reference conformation with respect to the torsional rotation, which will be denoted by θ. This reference conformation defines $\theta = 0°$ and must be known whenever employing the torsional potential function. Also, a direction of bond rotation must be selected and retained. Normally, the IUPAC convention provides the most straightforward means of defining $\theta = 0°$ and rotational direction. Unless the reference conformation, $\theta = 0°$, and rotation direction are known, it is not possible to unambiguously assign a torsional potential energy, $P(\theta)$, relative to a value of θ.

Both the quantum mechanical and the molecular mechanics energy of the model compound are determined as a function of θ at some incremental rotation $\Delta\theta$ ($\Delta\theta = 30°$ is a good working value). Bond lengths and angles are normally held constant. However, the effect of varying a bond length and/or angle on $P(\theta)$ can be discerned by simply repeating rotations in θ for different assigned valence geometries.

In some cases the quantum mechanical method will not converge because a $(1, 4)$ nonbonded interaction is sterically disallowed. Correspondingly, the molecular mechanics potential energy will be very large. These situations correspond to atom 1 and/or atom 4 not being protons. The solution is to make atom 1 and/or atom 4 proton(s). This will result in a model compound in which both the quantum and molecular mechanics energetics are reasonable at all θ, while introducing a minimal effect on the intrinsic orbital contributions to the torsional barrier about the bond. Moreover, the high energy state(s) will be realized in actual application since the origin(s) are in the nonbonded interactions(s).

Torsional parameterization can be achieved by the following steps:

(1) Compute the difference function, $P(\theta) = P_{QM}(\theta) - P_{MM}(\theta)$, as a function of θ according to the guidelines described above. $P_{QM}(\theta)$ is the quantum mechanical energy and $P_{MM}(\theta)$ is the molecular mechanics energy.

(2) Construct a plot of $P(\theta)$ vs. θ and identify locations of maxima and minima. Caution has to exercised in introducing additional fitting difficulties by retaining asymmetric extrema and/or "minor" extrema within, or about, "major" extrema. As a general rule, extrema less than one-half kcal/mol (height-maxima and depth-minima) should be neglected. These extrema make a minimal contribution to the fidelity of the calculations. They are usually due to differences in $(1, 3)$ and $(1, 4)$ interaction estimation between quantum and molecular mechanics, and greatly increase the difficulty of function fitting.

(3) Select the function(s) necessary to analytically reproduce $P(\theta)$. Caution, for the same reasons mentioned in step 2 above, should be exercised in attempting to reproduce potential shape asymmetry about extrema. In general, asymmetric well shape should be neglected.

At this point it becomes difficult to retain a general strategy because the procedures employed are dependent upon the mathematical function representation selected as well as specific $(\theta, P(\theta))$ set of values used in curve fitting. An Allinger force field[5] is very often used in free valence geometry molecular mechanics calculations. Hence curve fitting strategy will be discussed relative to the torsional functions used in this force field. A phase angle, α, has been added to each of the original torsional potential functions in order to be able to define an arbitrary rotation angle to have a zero torsional potential energy. The functions are

$$P_1(\theta) = \tfrac{1}{2} V_1 [1 + \cos(\theta - \alpha_1)] \qquad (7)$$

$$P_2(\theta) = \tfrac{1}{2} V_2 [1 - \cos(2\theta - \alpha_2)] \qquad (8)$$

$$P_3(\theta) = \tfrac{1}{2} V_3 [1 + \cos(3\theta - \alpha_3)] \qquad (9)$$

These three functions can be used singly, and in combinations, to reproduce a wide variety of multiple periodic extrema over a complete rotation. However, one commonly occurring torsional potential that cannot be constructed from equations (7)–(9) is a six-fold barrier. Thus this potential function constitutes an additional representation

$$P_4(\theta) = \tfrac{1}{2} V_4 [1 + \cos(6\theta - \alpha_4)] \qquad (10)$$

Plots of $P_i(\theta)$ vs. θ ($i = 1, 2, 3$) using $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ are shown in Figure 2 in order to

provide the reader a perspective for curve fitting. Orbital symmetry can be invoked to minimize torsional function fitting. This translates into assigning the periodicity and phase angle in the analytic function generated using eqs. (7)–(10), or combinations thereof (Fig. 2), by direct comparison to the computed $P(\theta)$.

Using this constraint the only unspecified set of variables are the function amplitude coefficients. The amplitude coefficients can be determined by a simple linear least-squares fit. For a linear combination of the torsional potential functions given in eqs. (7)–(9),

$$P(\theta) = P_1(\theta) + P_2(\theta) + P_3(\theta) \qquad (11)$$

the amplitude coefficients V_1, V_2, and V_3 are given by

$$V_j = \left\{ \left[\sum_i (1 + \cos\theta_i)P(\theta_i) \right] A_{j1} \right.$$
$$+ \left[\sum_i (1 - \cos 2\theta_i)P(\theta_i) \right] A_{j2}$$
$$\left. + \left[\sum_i (1 + \cos 3\theta_i)P(\theta_i) \right] A_{j3} \right\} / 2D$$
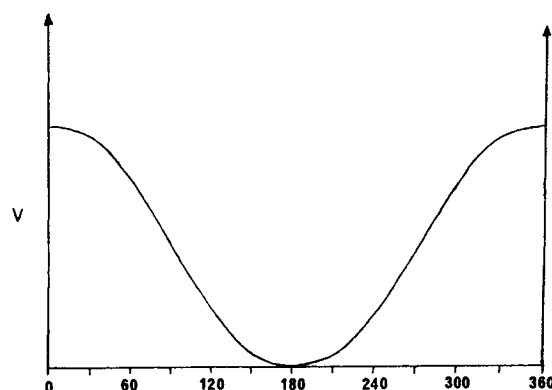
$$j = 1, 2, \text{and } 3 \qquad (12)$$

where

$$D = \tfrac{1}{4}\left[\sum_i (1 + \cos\theta_i)^2 A_{1,1} \right.$$
$$+ \sum_i (1 + \cos\theta_i)(1 - \cos 2\theta_i)A_{1,2}$$
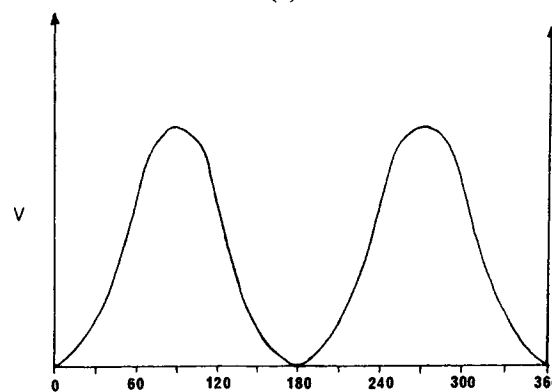$$\left. + \sum_i (1 + \cos\theta_i)(1 + \cos 3\theta_i)A_{1,3} \right] \qquad (13)$$

and

$$A_{1,1} = \sum_i (1 - \cos 2\theta_i)^2 \sum_i (1 + \cos 3\theta_i)^2$$
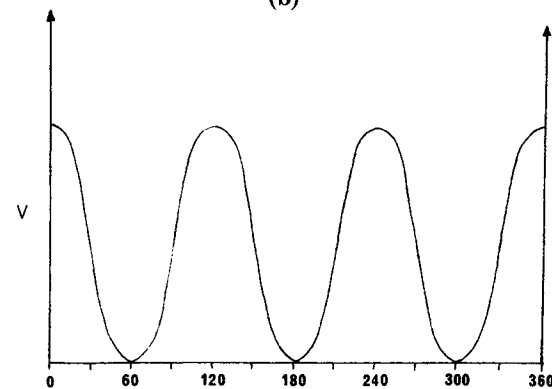$$- \left[\sum_i (1 - \cos 2\theta_i)(1 + \cos 3\theta_i) \right]^2 \qquad (14)$$

$$A_{1,2} = -\left[\sum_i (1 + \cos\theta_i)(1 - \cos 2\theta_i) \right.$$
$$\times \sum_i (1 + \cos 3\theta_i)^2$$
$$- \sum_i (1 - \cos 2\theta_i)(1 + \cos 3\theta_i)$$
$$\left. \times \sum_i (1 + \cos\theta_i)(1 + \cos 3\theta_i) \right] \qquad (15)$$

Θ, DEG.

(a)

Θ, DEG.

(b)

Θ, DEG.

(c)

Figure 2. $P(\theta)$ versus θ for the torsional potentials used in an Allinger force field [eqs. (7)–(9)].

$$A_{1,3} = \sum_i (1 + \cos\theta_i)(1 - \cos 2\theta_i)$$
$$\times \sum_i (1 - \cos 2\theta_i)(1 + \cos 3\theta_i)$$
$$- \sum_i (1 + \cos 2\theta_i)^2$$
$$\times \sum_i (1 + \cos\theta_i)(1 + \cos 3\theta_i) \qquad (16)$$

$$A_{2,2} = \sum_i (1 + \cos\theta_i)^2 \sum_i (1 + \cos 3\theta_i)^2$$

$$- \sum_i (1 + \cos\theta_i)(1 + \cos 3\theta_i)$$

$$\times \sum_i (1 + \cos\theta_i)(1 + \cos 3\theta_i) \qquad (17)$$

$$A_{2,3} = -\left[\sum_i (1 + \cos\theta_i)^2 \sum_i (1 - \cos 2\theta_i) \right.$$

$$\times (1 + \cos 3\theta_i) - \sum_i (1 - \cos 2\theta_i)^2$$

$$\left. \times \sum_i (1 + \cos\theta_i)(1 + \cos 3\theta_i) \right] \qquad (18)$$

$$A_{3,3} = \sum_i (1 + \cos\theta_i)^2 \sum_i (1 - \cos 2\theta_i)^2$$

$$- \left[\sum_i (1 + \cos\theta_i)(1 - \cos 2\theta_i) \right]^2 \qquad (19)$$

By symmetry $A_{2,1} = A_{1,2}$, $A_{3,1} = A_{1,3}$, and $A_{3,2} = A_{2,3}$. The index i refers to the observations (data points) realized from the molecular structure calculations. Appendix I contains a program listing that allows torsional parameter curvefitting based upon eqs. (12)–(19).

Cases 5 and 6 are examples torsion angle parameterizations carried out during the writing of this article. As expected, in case 5 rotation about the C—O bond leads to a two-fold potential barrier with minima for the carbonyl oxygen and hydroxyl proton cis and trans. The barrier heights of approximately 4.5 kcal/mol are much lower than the 17 kcal/mol estimated from mi-

crowave data for the C—O bond of formic acid.[16] The difference in barrier height is likely due to the structural differences between formic acid and the model compound.

Case 6 correctly identifies a high two-fold energy barrier to rotation about the C=C bond for both PCILO[17] and MNDO. The corresponding two minima occur for the cis and trans conformations about the bond. However, the heights of the torsional barriers for PCILO are much greater than the 50–60 kcal/mol barrier normally assigned to rotation about the double bond of ethylene.[18] The high barrier is not a consequence of any nonbonded interaction as evidenced by all the molecular mechanics potential energies being less than zero for each θ-dependent conformation employed. This is yet another example where a molecular orbital method can yield relationships between molecular geometry and molecular energetics at odds with experimental findings. However, as long as experimental molecular geometry is being sought, the PCILO barrier height inconsistency has minimal effect on the results. The model compound is only metastable and could tautomerize to $H(C=O)CH_2NH_2$. PCILO may not be parameterized to treat properly the orbital structure in the metastable compound.

This case is an example where more attention could be given to fitting well shapes about the two minima. The PCILO energetics are much too severe. Thus well shape fitting should be carried out using MNDO. Enhanced well shape fitting can be realized by increasing the number of data points near the minima and/or neglecting data points at, and around, barrier peaks.

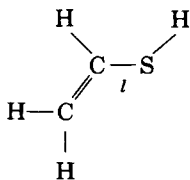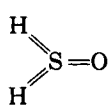**Table I.** Some examples of force-field parameterizations

Case 1	
A. Type, bond length	$C(sp^2)$—S(sulfide)
B. Model compound	
C. Bond lengths and angles	Taken from tables in the *Handbook of Chemistry & Physics*
D. Conformation	Planar conformation shown for the model compound
E. MO method	CNDO/2
F. MM method	Hopfinger steric parameters, Coulomb electrostatic with CNDO/2 charges

Table I (*continued*)

G. Raw curve fitting data

l (Å)	$(\Delta P)^a$(kcal/mol)
1.52	43.92
1.62	9.90
1.72	0.00
1.82	7.07
1.92	24.18
2.02	48.30
2.12	75.90

H. Parameters

k_l = 8.34 md/Å
l_0 = 1.76 Å
Q(C) = -0.0157
Q(S) = 0.0392
BM = -0.10 D

I. Experimental values

1. From x-ray analyses, l_0 = 1.80–1.83 Å (refs. 6–8)
2. From IR studies, k_l = 1.8–2.0 md/Å using
   1, 2-ethanedithiol (ref. 23)

Case 2

A. Type   bond length S(S—oxide)=O

B. Model compound

C. Bond lengths and angles   MNDO optimized structure
D. Conformation   no torsional bonds
E. MO method   MNDO and EHT
F. MM method   Hopfinger steric parameters,
                                 Coulomb electrostatic with MNDO and EHT,
                                 respectively, charges

G. Raw curvefitting data

l (Å)	[ΔP(MNDO)] (kcal/mol)	[ΔP(EHT)] (kcal/mol)
1.23	83.78	0.00
1.33	46.23	3.84
1.43	12.50	9.20
1.53	0.31	16.36
1.63	0.00	21.95
1.73	5.27	29.61
1.83	12.42	36.05

H. Parameters k_l = 7.29 md/Å
(from MNDO) l_0 = 1.64 Å
                         Q(S) = 0.228
                         Q(O) = -0.575
                         BM = 3.14 D

I. Experimental values

1. From x-ray analyses, l_0 = 1.45–1.49 (refs. 8, 12–14)
2. From microwave studies, k_l = 6–9 md/Å for a series of
   simple sulfones (ref. 24)
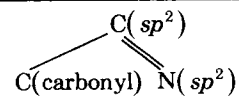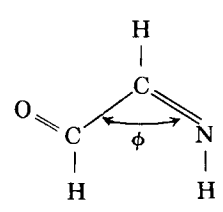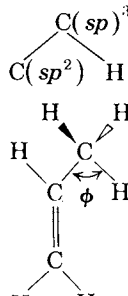
Case 3

A. Type, bond angle

B. Model compound

C. Bond lengths and angles   Taken from the *Handbook of
                                 Chemistry and Physics*
D. Conformation   Planar conformation as shown for the model
                                 compound

Table I (*continued*)

E. MO method	PCILO
F. MM method	Hopfinger steric potentials, Coulomb potential with PCILO charges, and a hydrogen-bond potential term

G. Raw curvefitting data

θ(degrees)	(ΔP) (kcal/mol)
118	1.98
120	1.03
122	0.20
124	0.04
126	0.00
128	0.22
130	0.74

H. Parameters $\quad k_0 = 1.73$ md Å/rad^2
$\theta_0 = 125.2°$

I. Experimental values
  1. From x-ray analyses $\theta_0 = 120\text{--}129°$ (refs. 8, 15)
  2. Not reported

Case 4

A. Type, bond angle

B. Model Compound

C. Bond lengths and angles	MINDO/3 optimized structure
D. Conformation	MINDO/3 optimized structure
E. MO method	MINDO/3
F. MM method	Scheraga steric potentials (ref. 25) Coulomb potential using MINDO/3 charges

G. Raw curvefitting data

θ (degrees)	(ΔP) (kcal/mol)
109.6	0.52
112.6	0.10
113.6	0.03
114.6	0.00
115.6	0.03
116.6	0.05
119.6	0.38

H. Parameters $\quad k_\phi = 0.81$ md Å/rad^2
$\theta_0 = 114.9°$

I. Experimental values
  1. From x-ray analyses, $\theta_0 = 110°\text{--}113°$ (ref. 15)
  2. From microwave studies, $k_\phi = 0.4\text{--}0.8$ md Å/rad^2 (ref. 26)

Case 5
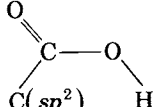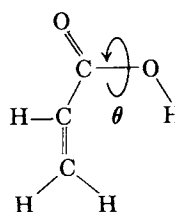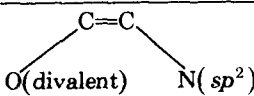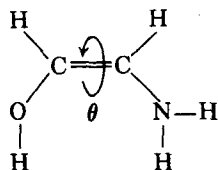
A. Type, torsion angle

B. Model compound

Table I (*continued*)

C. Bond lengths and angles	Taken from the *Handbook of Chemistry and Physics*
D. Conformation	$\theta = 0°$ defined by C—C—O—H cis planar. $\Delta\theta = 30°$ and minimum energy conformation for rotation about the C—C bond sought for each θ
E. MO method	CNDO/2
F. MM method	Hopfinger steric potentials, Coulombic potential with CNDO/2 charges, and a hydrogen bonding potential(ref. 2)

G. Raw curvefitting data

θ (degrees)	(ΔP) (kcal/mol)
0	0.98
30	1.82
60	4.23
90	4.61
120	3.90
150	1.73
180	0.00
210	1.61
240	376
270	4.39
300	4.20
330	1.77

H. Parameters[a]	$V_1 = 0.70$ kcal/mol, $\alpha_1 = 0°$ $V_2 = 4.49$ kcal/mol, $\alpha_2 = 0°$ $V_3 = 0.31$ kcal/mol, $\alpha_3 = 0°$
I. Experimental values	$V_2 = 17$ kcal/mol for formic acid using microwave spectroscopy (ref. 16)

Case 6

A. Type, torsional angle	
B. Model compound	
C. Bond lengths and angles	Taken from the *Handbook of Chemistry and Physics*
D. Conformation	*Cis* planar conformation defined as $\theta = 0°$ and $\Delta\theta = 30°$
E. MO method	PCILO and MNDO
F. MM method	Hopfinger steric potentials, Coulomb potential with PCILO and MNDO, respectively charges, and a hydrogen bonding potential

G. Raw curvefitting data

θ (degrees)	$[\Delta P(\text{PCILO})]$ (kcal/mol)	$[\Delta P(\text{MNDO})]$ (kcal/mol)
0	0.09	0.25
30	15.59	5.77
60	60.16	23.48
90	134.21	52.15
120	60.53	23.48
150	15.39	5.57
180	0.00	0.00
210	15.22	5.43
240	60.47	22.41
270	133.86	51.36
300	59.95	21.86
330	15.72	5.70

H. Parameters[a] (Using MNDO)	$V_1 = -2.74$ $V_2 = 43.01$ kcal/mol $V_3 = 2.60$
I. Experimental values	$V_2 = 50$–60 kcal/mol (ref. 18)

[a] In the Allinger force field (ref. 5) V_1, V_2, and V_3 are each divided by the number of atoms bonded to the two atoms forming the bond about which torsional rotation occurs.
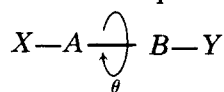
D. Quick Parameter Estimates

Circumstances can arise in which there is a need to get temporary rough estimates of force field parameters. For example, an investigator may plan to carry out a molecular orbital structure optimization calculation to obtain some electronic property and needs an initial trial structure. The trial structure could be generated using a molecular mechanics package if some missing parameters could be supplied. The time and effort of generating the missing parameters using the methods thus far described would be far beyond the target molecular orbital calculation in effort and complexity. Thus some simplified approach to force field parameter estimation is needed.

Methods to make rough force-field parameter estimates can again be divided into two categories: (a) bond lengths and bond angles and (b) torsional rotations. The l_0 and ϕ_0 for bond lengths and angles can be taken from literature data bases including the *Handbook of Chemistry and Physics*, the *Chemist's Companion*, and the *Cambridge X-ray Crystallographic Data Base*. The force constants k_l and k_ϕ should each be set equal to a relatively large number. $k_l = 10$ md/A and $k_\phi = 2.5$ md A/rad^2 are suggested values. This will insure that the energy optimized structure retains bond lengths and angles quite close to l_0 and ϕ_0. The potential pitfall of this approach is to unrealistically distort other bond lengths and/or angles that are "correctly" parameterized, but must unrealistically deform due to these artificially assigned high values of k_l and k_ϕ.
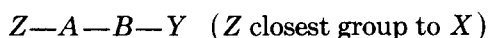
The most straightforward way of treating bond moments in a quick and rough parameterization is to set each of them equal to zero. If the investigator is able to carryout a Del Re charge calculation,[19] or other method dependent only upon bonding topology, then the bond moments can be quickly estimated using the resultant charges in eq. (6).

The most general rule for quick and rough torsional potential parameterization is as follows:
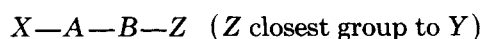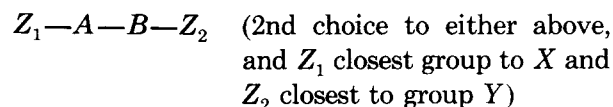
If the torsional rotation in question is

$$X-A\overset{\frown}{\underset{\theta}{\smile}}B-Y$$

and torsional potential parameters for

$$Z-A-B-Y \quad (Z \text{ closest group to } X)$$

or

$$X-A-B-Z \quad (Z \text{ closest group to } Y)$$

or

$$Z_1-A-B-Z_2 \quad \text{(2nd choice to either above, and } Z_1 \text{ closest group to } X \text{ and } Z_2 \text{ closest to group } Y)$$

are available, then the known torsional potential parameters for the similar system is assigned to $X-A-B-Y$. The term "closest" refers to selecting an atom, or atomic group, having bonding symmetry and size as near as possible to X and/or Y. If $X-A$ and/or $B-Y$ is a double bond, then the selected similar system must contain this/these double bonds. Another simple rule that pertains to parameterizing double bonds is to use eq. (8) with $V_2 = 15$ kcal/mol. When no parameterized torsional bond containing $-A$ and/or $-B-$ is available to model the $X-A-B-Y$ structure, then parameterized systems of the form

$$Z-A'-B'-Y$$

or

$$X-A'-B'-Z$$

or

$$Z_1-A'-B'-Z_2$$
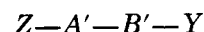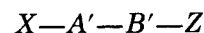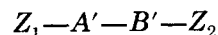
where A' and B' have the same bonding symmetry to A and B, respectively, should be sought and used.

DISCUSSION

The procedures and methods reported here for force-field parameterization are premised upon restricted applications to problems of molecular geometry. Heats of formation, vibrational frequencies, heat capacities, and other molecular properties not solely a component of molecular architecture cannot be expected to be reliably estimated. The parameterized molecular mechanics force field will be approximately of the same "quality" in estimating molecular geometry as the molecular orbital method used in the parameterization process. Implicit in making this statement is the assumption that a single quantum mechanical procedure and a single molecular mechanics method have been used in all the parameter calculations.

There is no problem in maintaining self-consistency in use of the molecular mechanics method. However, as evident from the case study parameterizations reported in Table I and Figure 1, the molecular orbital methods can fail to give

realistic results depending upon the molecular structure under consideration. Thus the self-consistency rule for molecular orbital methods application must be modulated to take into account methodology breakdown. Guidelines to follow when a quantum method fails are to (1) lookup as much experimental data on the model compound as possible, (2) make a series of chemical substitutions in the model compound including methyl substitutions for protons and/or halogenations for protons, and (3) carry out parameterization studies on the compounds in step 2 using all available molecular orbital methods. The primary goal in applying these guidelines is to arrive at a reasonable set of force-field parameters. The secondary goal is to identify the chemical structure in the model compound which is responsible for the breakdown of the quantum method(s). The investigator may then be in a position to decide upon an alternate strategy to generate the parameters.

This discussion raises the question of which molecular orbital and molecular mechanics methods are best to use. In the case of the nonbonded molecular mechanics potentials, the choice in molecular mechanics method has minimal effect on parameterization reliability and, consequently, the issue is self-consistency. Hence the molecular mechanics method used by an investigator in general applications research should also be the one used in parameterization experiments.

The best choice in quantum mechanics methodology for force-field parameter generation cannot be stated. In theory *ab initio* quantum mechanical methods should give the most reliable and accurate force-field parameters. Perhaps, a consensus will ultimately be reached among computational chemists to develop a "universal" molecular mechanics force field. An *ab initio* quantum mechanical formalism would likely be selected for use in the parameterization process. The obvious drawback to employing *ab initio* methods is the huge expenditure of computational time and effort to generate parameters. However, not so obvious, but nevertheless a consideration, is that the molecular mechanics force-field potential representations currently used may not justify the deployment of sophisticated quantum mechanical calculations. For example, the treatment of bond length potential energy by decoupling chemical bonds in molecular mechanics formalisms negates application of detailed resonance behavior realized from quantum calculations.

In our laboratories, largely by process of evolution, the MINDO/3 method[20] has become the principal quantum mechanics method used in molecular mechanics parameterizations. For model compounds involving hydrogen bonding, the MINDO/3H[21] (MINDO/3 corrected to treat hydrogen bonds) is employed whenever possible and questionable parameterizations are backed up with subsidiary calculations using CNDO/2[22] and PCILO.

Lastly, the "quick and rough" parameterization procedures should not be used as general tools to fill in molecular mechanics software force field parameter tables. These quick rules are only intended to aid in the generation of trial structures for additional optimization calculations.

APPENDIX I

A BASIC program to generate bond length, bond angle, and torsional potential force field parameters. The potential functions and geometric and energy units employed are listed in the program.

```
1 DIM A(3,3),P(36),TH(36)
5 PRINT"THIS PROGRAM COMPUTES THE FORCE CONSTANTS FOR THE"
6 PRINT"FUNCTIONS -->
7 PRINT"BOND LENGTH P(L)  = K(L)*(L-L(0))**2"
8 PRINT"BOND ANGLE   P(PHI) = K(PHI)*(PHI - PHI(0))**2"
9 PRINT"TORSIONAL ANGLE P(THETA)=.5*V(1)*(1+COS(THETA))+"
10 PRINT".5*V(2)*(1-COS(2*THETA))+.5*V(3)*(1+COS(3*THETA))"
12 PRINT "THE UNITS ASSUMED ARE 1) LENGTH = ANGSTROMS,"
13 PRINT "2) ANGLES = DEGREES, AND 3) ENERGY = KCAL/MOLE"
15 PRINT"  "
20 PRINT"*************************************************"
22 PRINT"INPUT NUMBER OF DATA POINTS -->":INPUT N
25 PRINT"INPUT 0 TO COMPUTE TORSIONAL POTENTIALS"
26 PRINT"INPUT 1 TO COMPUTE BOND LENGTHS"
27 PRINT"INPUT 2 TO COMPUTE BOND ANGLES"
30 PRINT"INPUT CHOICE -->": INPUT IC
32 PL=99999.
40 IF IC = 0 THEN 2000
45 FOR I = 1 TO N
48 IF IC >1 THEN 52
50 PRINT "INPUT BL";I;"AND POT. ENERGY";I:INPUT TH(I),P(I)
51 IF IC = 1 THEN 60
52 PRINT"INPUT BA";I;"AND POT. ENERGY";I:INPUT TH(I),P(I)
60 IF P(I)>PL THEN 70
65 PL=P(I):VL=TH(I)
70 NEXT I
80 FOR I = 1 TO N
85 PRINT""
90 TH(I) = TH(I) -VL: P(I)=P(I)-PL
95 NEXT I
100 PRINT"THE RAW DATA TRANSFORMED TO THE POTENTIAL"
110 PRINT"ENERGY MINIUM POSITION IS"
120 PRINT"*************************************************"
130 PRINT"  "
140 PRINT " INDEPENDENT VARIABLE      POTENTIAL ENERGY"
1000 FOR I = 1 TO N
1005 PRINT"        ";TH(I);"              ";P(I)
1010 PRINT"  "
1012 NEXT I
1015 A4=0:A3=0:A2=0:A1=0:AN=0:B2=0:B1=0:B=0
1020 PRINT"  "
1030 FOR I = 1 TO N
1040 A4=A4+TH(I)[4:A3=A3+TH(I)[3:A2=A2+TH(I)[2
1050 A1=A1+TH(I):AN=AN+1:B2=B2+P(I)*TH(I)[2
1060 B1=B1+P(I)*TH(I):B=B+P(I)
1070 NEXT I
1080 A(1,1)=AN*A2-A1*A1:A(1,2)=-(AN*A3-A2*A1)
1090 A(1,3)=A3*A1-A2*A2:A(2,1)=-(AN*A3-A2*A1)
1100 A(2,2)=AN*A4-A2*A2:A(2,3)=-(A4*A1-A3*A2)
```

```
1110 A(3,1)=A3*A1-A2*A2:A(3,2)=-(A4*A1-A3*A2)
1120 A(3,3)=A4*A2-A3*A3
1130 D=A4*(A2*AN-A1*A1)-A3*(A3*AN-A2*A1)
1140 D = D + A2*(A3*A1-A2*A2)
1150 PRINT" "
1155 PRINT " "
1160 AK=(B2*A(1,1)+B1*A(1,2)+B*A(1,3))/D
1170 PRINT""
1175 PRINT""
1180 CH=(A(2,1)*B2+A(2,2)*B1+A(2,3)*B)/D
1185 AL=CH/(-2*AK)
1190 AL=AL+VL
1200 IF IC>1 THEN 1250
1210 AX=AK/71.94
1220 PRINT "BOND LENGTH FORCE CONSTANTS"
1230 PRINT "L(0) = ";AL;" ANGSTROMS"
1240 PRINT "K(L) = ";AK;" KCAL/ANGSTROM**2"
1245 PRINT "K(L) = ";AX;" MD/ANGSTROM"
1248 IF IC <2 THEN 5
1250 PRINT "BOND ANGLE FORCE CONSTANTS"
1260 PRINT "PHI(0) = ";AL;" DEGREES"
1270 AX = AK/.021914
1280 PRINT "K(PHI) = ";AK;" KCAL/DEGREE**2"
1290 PRINT "K(PHI) = ";AX;" MD-ANGSTROM/RAD**2"
1300 IF IC = 2 THEN 5
2000 A1=0:A2=0:A3=0:A4=0:A5=0:A6=0:A7=0:A8=0:A9=0
2010 PRINT"BE SURE THAT YOUR THETA RANGE IS 360 DEG" .
2030 FOR I = 1 TO N
2040 PRINT "INPUT THETA(";I;") AND P(THETA";I;")"
2050 INPUT TH(I),P(I)
2060 TH(I)=TH(I)*.0174533
2070 A1=A1+(1+COS(TH(I)))[2:A2=A2+(1-COS(2*TH(I)))[2
2080 A3=A3+(1+COS(3*TH(I)))[2
2090 A4=A4+(1+COS(TH(I)))*(1+COS(3*TH(I)))
2100 A5=A5+(1+COS(TH(I)))*(1-COS(2*TH(I)))
2110 A6=A6+(1-COS(2*TH(I)))*(1+COS(3*TH(I)))
2120 A7=A7+(1+COS(TH(I)))*P(I):A8=A8+(1-COS(2*TH(I)))*P(I)
2130 A9= A9+(1+COS(3*TH(I)))*P(I)
2140 NEXT I
2150 A(1,1)=A2*A3-A6*A6:A(1,2)=-(A5*A3-A4*A6)
2160 A(1,3)=A5*A6-A2*A4:A(2,1)=A(1,2)
2170 A(2,2)=A1*A3-A4*A4:A(2,3)=-(A1*A6-A4*A5)
2180 A(3,1)=A(1,3):A(3,2)=A(2,3)
2190 A(3,3)=A1*A2-A5*A5
2200 D=(A1*A(1,1)+A5*A(1,2)+A4*A(1,3))/4.
2210 V1=A7*A(1,1)+A8*A(1,2)+A9*A(1,3)
2220 V2=A7*A(2,1)+A8*A(2,2)+A9*A(2,3)
2230 V3=A7*A(3,1)+A8*A(3,2)+A9*A(3,3)
2240 V1=V1/(2*D):V2=V2/(2*D):V3=V3/(2*D)
2250 PRINT"THE TORSIONAL CONSTANTS ARE (IN KCAL/MOLE)
2260 PRINT"V(1) = ";V1:PRINT"V(2) = ";V2:PRINT"V(3) = ";V3
2270 IF IC=0 THEN 5
```

References

1. A. Warshel, in *Semiempirical Methods of Electronic Structure Calculation*, G. A. Segal, Ed., Plenum, New York, 1977, p. 133.
2. A. J. Hopfinger, *Conformational Properties of Macromolecules*, Academic, New York, 1973.
3. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, 4, 187 (1983), and references therein.
4. S. Karplus and S. Lifson, *Biopolymers*, 10, 1973 (1971).
5. (a) For a detailed reference listing, see E. Osawa in *Quantum Chemistry Program Exchange Bull.*, 3, 87

(1983); (b) N. L. Allinger, *J. Am. Chem. Soc.*, 99, 8127 (1977), and references therein; (c) N. L. Allinger, *Quantum Chemistry Program Exchange*, 395, 410 (1977).
6. *Handbook of Chemistry and Physics*, Chemical Rubber, Cleveland, OH, 1984.
7. *The Chemist's Companion*, A. J. Gordon and R. A. Ford, Eds., Wiley, New York, 1972.
8. *Cambridge X-ray Crystallographic Data Base*, O. Kennard, Ed., University Chemical Laboratory, Cambridge, England.
9. T. A. Ryan, Jr., B. L. Joiner, and B. F. Ryan, *MINITAB Student Handbook*, Duxbury, North Scituate, MA, 1976.
10. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, 99, 4899, 4907 (1977).
11. R. Hoffman, *J. Chem. Phys.*, 21, 466, 767 (1953).
12. P. H. Lauer, in *Sulfur in Organic and Inorganic Chemistry*, A. Senning, Ed., Dekker, New York, 1972, Vol. 3.
13. C. C. Price and S. Oae, *Sulfur Bonding*, Ronald, New York, 1962.
14. N. Kharasch, *Organic Sulfur Compounds*, Pergamon, New York, 1961, Vol. 1.
15. O. Kennard, D. G. Watson, F. H. Allen, N. W. Isaacs, W. D. S. Motherwell, R. C. Pettersen, and W. G. Town, *Molecular Structures and Dimensions*, University Chemical Laboratory, Cambridge, England.
16. M. Simonetta and S. Carra, in *The Chemistry of Carboxylic Acids and Esters*, S. Patai, Ed., Wiley, New York, 1969, p. 3.
17. S. Diner, S. P. Malriew, F. Jordan, and M. Gilliert, *Theor. Chim. Acta*, 15, 100 (1969).
18. J. E. Douglas, B. S. Rabinowitch, and S. F. Looney, *J. Chem. Phys.*, 23, 315 (1955).
19. G. Del Re, B. Pullman, and T. Yonezawa, *Biochim. Biophys. Acta*, 75, 1285 (1975).
20. R. C. Bingham, M. J. S. Dewar, and D. H. Lo, *J. Am. Chem. Soc.*, 97, 1285.
21. S. N. Mohammad and A. J. Hopfinger, *Int. J. Quantum Chem.*, 22, 1189 (1982).
22. J. A. Pople, D. P. Santry, and G. A. Segal, *J. Chem. Phys.*, 43, 5129 (1965); J. A. Pople and G. A. Segal, *ibid.*, 43, 5136 (1965); 44, 3289 (1966).
23. M. Hayashi, Y. Shiro, T. Oshima, and H. Murata, *Bull. Chem. Soc. Jpn.*, 38, 1734 (1965).
24. J. Brunvoll and I. Hargittai, *Acta Chim. Acad. Sci. Hung*, 96, 337 (1978).
25. R. A. Scott and H. A. Scheraga, *J. Chem. Phys.*, 45, 2091 (1966).
26. A. Bouchy and G. Rolussy, *J. Chim. Phys. Phys. Chim. Biol.*, 78, 669 (1981).