Kim Palmo
Berit Mannfors
Noemi G. Mirkin
Samuel Krimm
*Biophysics Research Division
and Department of Physics,
University of Michigan,
Ann Arbor, MI 48109*

Potential Energy Functions: From Consistent Force Fields to Spectroscopically Determined Polarizable Force Fields

Abstract: *We review our methodology for producing physically accurate potential energy functions, particularly relevant in the context of Lifson's goal of including frequency agreement as one of the criteria of a self-consistent force field. Our spectroscopically determined force field (SDFF) procedure guarantees such agreement by imposing it as an initial constraint on parameter optimization, and accomplishes this by an analytical transformation of ab initio "data" into the energy function format. After describing the elements of the SDFF protocol, we indicate its implementation to date and then discuss recent advances in our representation of the force field, in particular those required to produce an SDFF for the peptide group.* © 2003 Wiley Periodicals, Inc. Biopolymers 68: 383–394, 2003

Keywords: *molecular mechanics energy function; polarizability; vibrational frequencies; N-methylacetamide; charge fluxes; ir intensities; peptide torsion; amide NH out-of-plane bend*

INTRODUCTION

The importance of computational simulations of macromolecular structure, energetics, and dynamics in enhancing our understanding of physical properties and biological behavior of such molecules is an accepted paradigm in current scientific research. There is also a growing awareness that the success of this endeavor hinges on the ability of the potential energy functions used in the simulations to accurately reproduce the physics of these systems. As recently commented on in regard to protein structure prediction, "Despite many years of development of molecular simulation methods, attempts to refine models that are already relatively close to the native structure have met with relatively little success. The failure is likely to be due to inaccuracies in the potential functions used in the simulations . . ."[1] It is not surprising, therefore, that much effort has been devoted to achieving improved accuracy in such energy functions.

Shneior Lifson was instrumental in conceptualizing our modern approach to producing more accurate so-called molecular mechanics (MM) energy functions, doing so by requiring that they reproduce a complete range of molecular properties and by utiliz-

Biopolymers, Vol. 68, 383–394 (2003)

ing least-squares optimization of energy parameters to the experimental data on these properties. In a seminal series of papers on what he termed a consistent force field (CFF),[2–5] he demonstrated how such an approach led to a significant improvement in the predictive capabilities of energy functions.

One of the properties Lifson gave particular attention to in order to produce an accurate potential energy surface was vibrational frequencies. Good structure agreement is necessary but it is not a sufficient condition for good reproduction of vibrational motion. Frequencies are very sensitive to the potential function, and since they are also one of the experimentally most accurately determined properties of molecules, they are very suited for use as a measure of the quality of an energy function. Although Lifson's early potentials[2–4] led to RMS errors (for non-CH stretch frequencies) in the range of 25–40 cm^{-1}, rendering their spectroscopic reliability inherently suspect, subsequent efforts,[5,6] by introducing cross-terms in the bonded part of the potential, improved on this, in a few cases[6] bringing the RMS error into the respectable 5–10 cm^{-1} range.

Many other efforts have extended and developed the CFF approach in the succeeding years, but the accurate prediction of vibrational frequencies and modes has not been their primary concern. As a result, current standard MM functions at best give RMS errors in about the same range as those of the early Lifson work. We have felt that such inadequacy inherently restricts the ultimate physical accuracy of such functions, and have therefore sought to incorporate frequency agreement as an added initial constraint in the optimization of energy function parameters. We call such a function a spectroscopically determined force field (SDFF), and find that requiring spectroscopic accuracy can also be a means of sensitively detecting needed fundamental physical components of the force field. It should be emphasized that the SDFF gives as good, if not better, structures and energies as current standard force fields, and provides far superior reproduction of vibrational frequencies. The quality of the energy function is also crucial for molecular dynamics trajectories, especially in simulations (such as protein–drug docking calculations) that rely on accurate molecule–molecule and intramolecular interactions.

In this paper we review the philosophy and methodology of the SDFF approach, describe its implementation to date on hydrocarbon systems, and briefly indicate new directions involved in extending these procedures to polar systems such as polypeptides.

SPECTROSCOPICALLY DETERMINED FORCE FIELDS

SDFF Philosophy

As mentioned above, our goal is to ensure reliable reproduction of vibrational frequencies (and eigenvectors), in addition to structures and energies, in the optimization of the MM parameters. Rather than doing this by a least-squares optimization of force constants, we have used an analytical transformation that determines the valence parameters from ab initio (i.e., quantum mechanical, QM) equilibrium structures, and Hessians, in combination with given MM nonbonded potential energies.[7,8] If necessary, the ab initio force constants can be scaled to experimental frequencies and band assignments. The QM calculations also provide other useful "data" on the molecule. A significant advantage of using QM results is that consistent data of uniform quality are obtained for a given class of molecules. Using QM data in the construction of energy functions favors intramolecular and small cluster properties over condensed phase observables, but there need be no serious conflict regarding reproduction of such experimental data, especially if many-body interactions are properly included.

Since the SDFF procedure depends on initial knowledge of the nonbonded potential (see below), careful attention must be given to the form of this potential and to the protocol for determining the associated MM parameters. The analytical transformation can guarantee frequency agreement, but this may be at the expense of nonsensical force constants if these are uncritically taken as the average of the values in the different molecules and conformations. This would seriously limit the range of properties that can transferably be reproduced. However, inconsistencies in the valence parameters are easy to detect in the SDFF transformations and are mostly traceable to a physically inadequate description of the nonbonded interactions. This sensitivity is helpful in developing more transferable models. We expand on these issues after discussing the specific features of the SDFF methodology.

SDFF Methodology

There are four elements in the SDFF methodology. First, a form is selected for the MM potential, one that is (hopefully) inclusive enough to incorporate all the physically important contributions. In the first implementation of this approach, viz., for linear saturated hydrocarbon chains,[9] we found that a standard potential that contained more cross-terms than typically

used (including internal coordinate/torsion and torsion/torsion) was sufficient to give a satisfactory accounting of vibrational frequencies as well as of structures and energies. Because of the importance of taking into account changes in the electron distribution with geometry and conformational changes, our current SDFF includes atomic dipole as well as charge and dipole flux terms and explicitly includes anharmonic contributions (which are parametrized directly from the model molecules studied; see below). Such a form of the potential is given by

$$V = \tfrac{1}{2} \sum_i f_{ii}(r_i - r_{i0})^2 + \tfrac{1}{2} \sum_i c_{i1}(r_i - r_{i0})^3$$
$$+ \tfrac{1}{2} \sum_i c_{i2}(r_i - r_{i0})^4 + \sum_{i<j} f_{ij}(r_i - r_{i0})(r_j - r_{j0})$$
$$+ \sum V_{\text{tor}} + \sum V_{r,\text{tor}} + \sum V_{\text{tor,tor}} + \sum V_{\text{nb}} \quad (1)$$

where the r_i are internal coordinates whose intrinsic reference values are r_{i0}; the f_{ii}, c_{i1}, and c_{i2} are the quadratic, cubic, and quartic valence force constants; and the f_{ij} are interaction force constants for coordinates for which the potential energy has an approximately quadratic dependence (the valence parameters also include a possible conformation dependence). The torsion potential is represented by a Fourier cosine series,

$$V_{\text{tor}} = \sum_m V_m(1 \pm \cos m\chi) \quad (2)$$

where V_m is the barrier to rotation of periodicity m associated with torsion angle χ. The MM torsion coordinate for a bond can be defined as a linear combination of some or all of the individual torsion coordinates associated with the bond. In order to guarantee orthogonality with out-of-plane coordinates associated with sp^2 bonds, we use Bell's torsion,[5,10] i.e., $\chi = \tfrac{1}{2}(\chi_1 + \chi_2)$, where χ_1 and χ_2 are dihedral angles that share the central bond but have different arms. This definition, which for C=C bonds actually measures the angle between the π orbitals, is crucial to properly representing the torsion potential in alkenes[11] (see below). The cross-terms that involve torsions are given by

$$V_{r,\text{tor}} = f_{ij}(r_i - r_{i0})g(\chi_j) \quad (3)$$

if only one of the interacting coordinates is a torsion, and by

$$V_{\text{tor,tor}} = f_{ij}g(\chi_i)g(\chi_j) \quad (4)$$

if both coordinates are torsions. The function $g(\chi)$ is defined by

$$g(\chi) = \frac{1}{m} \sin m(\chi - \chi_0) \quad (5)$$

where χ_0 is the nearest intrinsic minimum of the torsion potential and m is the periodicity of the dominant term. The form of $g(\chi)$, obtained from the conformation dependence of ab initio interaction force constants, makes $V_{r,\text{tor}}$ and $V_{\text{tor,tor}}$ behave like standard (quadratic) cross-terms when the torsion coordinates are close to their intrinsic minima, and properly accounts for the periodicity elsewhere. This includes a sign reversal of the interaction force constant for torsion deformations larger than $\pm\pi/(2m)$.

The nonbonded potential accounts for van der Waals and electrostatic interactions by

$$V_{\text{nb}} = \frac{A_i A_j}{R_{ij}^9} - \frac{B_i B_j}{R_{ij}^6} + \frac{1}{4\pi\kappa\varepsilon_0}\left[\frac{q_i q_j}{R_{ij}} + \frac{q_i \boldsymbol{\mu}_j \cdot \mathbf{R}_{ij}}{R_{ij}^3} \right.$$
$$\left. + \frac{R_{ij}^2 \boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j - 3(\mathbf{R}_{ij} \cdot \boldsymbol{\mu}_i)(\mathbf{R}_{ij} \cdot \boldsymbol{\mu}_j)}{R_{ij}^5} \right] \quad (6)$$

where $\mathbf{R}_{ij}$ is a vector from atom i to atom j, q_i is the charge on atom i, $\boldsymbol{\mu}_i$ is the atomic dipole on atom i, ε_0 is the permittivity of free space, κ is the dielectric constant, and A_i and B_i are respectively the repulsive and attractive van der Waals parameters of atom i (we have found that the 9-6 form gives better results than the 12-6 form). The nonbonded potential energy is summed over all atom pairs in positions 1,4 and higher. We do not include separate terms for hydrogen bonds since we have shown that their contribution is satisfactorily accounted for by a new formulation of the relevant electrical interactions.[12]

The atomic charges, q_i, are taken as the sum of bond charge increments (BCIs), $q_i(b)$,

$$q_i = \sum_b q_i(b) \quad (7)$$

where the index b runs over all bonds that contain atom i. The BCIs can depend on the internal coordinate r_j, i.e.,

$$q_i(b) = q_{i0}(b) + \sum_j a_{bj}(r_j - r_{j0}) \quad (8)$$

where $q_{i0}(b)$ is the BCI when all $r_j = r_{j0}$ (the intrinsic equilibrium geometry) and a_{bj} is the charge flux along bond b (attached to atom i) with respect to internal coordinate r_j. For charge fluxes associated with torsion coordinates, $(r_j - r_{j0})$ is replaced by $\cos \chi_j$.

By analogy, atomic dipoles are described by

$$\boldsymbol{\mu}_{i0} = \sum_b \mu_i(b)\mathbf{e}_b \tag{9}$$

where $\mu_i(b)$ is the dipole increment of atom i along bond b and $\mathbf{e}_b$ is a unit vector along bond b, and the possible geometry dependence is given by

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_{i0} + \sum_j \mathbf{d}_{ij}(r_j - r_{j0}) \tag{10}$$

where $\mathbf{d}_{ij}$ is the dipole flux on atom i with respect to coordinate r_j.

Polarization is not a critical component of the SDFFs for the relatively nonpolar hydrocarbons,[9,11,13] but it cannot be neglected for polar groups like the peptides. We have therefore sought to implement polarization, and have done so[14] as a selective combination of induced BCIs and anisotropically induced atomic dipoles. Thus, under the influence of an electric field, an induced charge

$$q_{i,\text{ind}} = \sum_b 4\pi\varepsilon_0\alpha(b)E_p(b)/r_b \tag{11}$$

is added to the existing charge on atom i. Here the index b runs over the bonds containing atom i, $\alpha(b)$ is the bond polarizability, $E_p(b)$ is the magnitude of the electric field parallel to the bond (taken as the average of the fields at the atoms of the bond), and r_b is the length of the bond b. The two atoms of a bond get induced charges that are equal in magnitude but opposite in sign. Similarly, an induced atomic dipole,

$$\boldsymbol{\mu}_{i,\text{ind}} = \sum_b 4\pi\varepsilon_0[\alpha_{\text{p},i}(b)\mathbf{E}_{\text{p},i}(b) + \alpha_{\text{pp},i}(b)\mathbf{E}_{\text{pp},i}(b)] \tag{12}$$

is added to the existing dipole on atom i. $\mathbf{E}_{\text{p},i}(b)$ and $\mathbf{E}_{\text{pp},i}(b)$ are, respectively, the parallel and perpendicular (with respect to the bond b) electric fields at atom i, and $\alpha_{\text{p},i}(b)$ and $\alpha_{\text{pp},i}(b)$ are the parallel and perpendicular (with respect to the bond b) atomic polarizabilities, respectively. This bond increment approach to achieve anisotropic atomic polarizability, which was first used by us for water and formaldehyde,[14] has since been applied successfully to various organic compounds.[15] Since the electric field depends

on the induced charges and dipoles, and vice versa, an iterative procedure is needed to compute these quantities. A procedure is also required to avoid the "polarization catastrophe" resulting from interactions between too closely interacting induced dipoles.[16]

The second element in the SDFF methodology is the ability to make a direct transformation from the spectroscopic to the MM force field. We have shown[8] that this is achieved through an analytical expression for the force constants in terms of the ab initio Hessians, $\partial^2 V/\partial x_\alpha \partial x_\beta$, and the nonquadratic potential V_{nb} (the quadratic part includes the valence and torsion terms)

$$f_{ij} = \sum_\alpha \left[-Q_{\alpha,ij}\frac{\partial V_{\text{nb}}}{\partial x_\alpha} + \sum_\beta P_{\alpha i}P_{\beta j}\left(\frac{\partial^2 V}{\partial x_\alpha \partial x_\beta} - \frac{\partial^2 V_{\text{nb}}}{\partial x_\alpha \partial x_\beta}\right) \right] \tag{13}$$

where

$$P_{\alpha i} = \frac{\partial x_\alpha}{\partial r_i} \quad \text{and} \quad Q_{\alpha,ij} = \frac{\partial^2 x_\alpha}{\partial r_i \partial r_j}$$

The intrinsic geometry parameters r_{i0} are then given by the solution of the system of linear equations,

$$\sum_j f_{ij}(r_j - r_{j0}) = -\sum_\alpha \frac{\partial V_{\text{nb}}}{\partial x_i}P_{\alpha i}, \quad i = 1, \ldots, N_r \tag{14}$$

where N_r is the number of internal coordinates. Since this transformation is analytic, it preserves in the SDFF the exact frequencies and structure of the scaled ab initio calculation. Although the transformation is initiated by assuming a starting set of V_{nb} parameters (we discuss below the determination of the initial values), these can be subsequently optimized (fine tuned) in the refinement procedure, if necessary.[17]

The third element is the application of the SDFF transformation to the ab initio structures and Hessians of a set of conformers of the model molecules for the macromolecular system. This also allows us to determine specific analytic forms of the possible conformation dependence of the valence terms. In the case of the saturated hydrocarbons, the set consisted of the 14 stable conformers of *n*-pentane and *n*-hexane[9,18] plus the 7 stable conformers of isopentane, 3-methylpentane, and neopentane.[13] In the case of the olefinic hydrocarbons,[11] this consisted of 13 structures of ethene, propene, *skew* and *syn*-1-butene, *trans*- and *cis*-2-butene, and isobutene and some of their deuterated species. The nonbonded parameters are then optimized under the standard MM assumption that the

intrinsic force constants (f_{ii}, f_{ij}) and *intrinsic* geometry parameters (r_{i0}) are the same for all conformers. The relative (ab initio) conformer barriers and energies are incorporated in optimizing the torsion parameters V_m.

In the final element of the procedure the set of force constants is systematically reduced, based on a *predetermined limit* on the frequency error,[8] since it is neither desirable nor necessary to include in *V* the huge number of f_{ij}, most of which are very small and do not significantly affect the vibrational frequencies. All of these operations, as well as structure optimization, normal mode analysis, molecular dynamics, etc., are implemented in SPEAR, our molecular modeling package.[19]

Optimization Protocol and Nonbonded Potential

The order in which parameters are determined is important, in that correlations usually occur between them. In order to minimize these, we derive the parameters for a chosen set of model molecules in the following order: electrostatic, van der Waals (vdW), valence, and torsion. The electrostatic parameters are determined first because they can be optimized to QM electric potentials independently of all other parameters (other types do not contribute to the electric potential). Initial (but very close to final) vdW parameters are then determined by fitting to QM potential energy data of dimers or complexes of the model molecules. When the nonbonded parameters are known, the SDFF transformation[8] is used to compute valence parameters (other than torsions) by analytically transforming QM minimum energy structures and (scaled) force constants into MM force constants and intrinsic geometry parameters, simultaneously fine-tuning the vdW parameters.[17] Next, the torsion potentials are optimized using QM torsion barrier data. Finally, indeterminacy in the valence force constants, caused by local redundancies, is removed.[20] This makes it possible to slightly reoptimize (some of) the force constants to compensate for the averaging of parameters in the different conformers and for the effects of the f_{ij} that were dropped. This sequence minimizes correlations between energy parameters and guarantees the best V_{nb} for the transformations of Eqs. (13) and (14).

Electrostatic Parameters. The electrostatic parameters are determined by fitting to QM electric potentials. Mathematically, the optimization to the electric potential is a linear (or when polarization is used, "almost linear") least-squares fit. Thus, initial values

for the parameters do not have to be supplied, and there is no multiple minimum problem in parameter space. Our method for optimizing the parameters to the electric potential[14] is similar to other recently published methods,[21–23] but there are some important differences. The electrostatic interactions are of long range, and in order to avoid having the parameters determined solely from grid points closest to the atoms (where the electric potential is strongest), we use a weighting factor for each point equal to the distance to the nearest atom. Our method is simpler than that of others[22] because no probe molecule and no vdW potential have to be assumed. This is particularly important in our approach, since the electrostatic parameters are the first ones to be determined in a new force field.

Instead of using one of the generic schemes for generating grid points, we use GAMESS[24] to generate the points and to compute the electric potential on carefully chosen planes near or through the molecules. The planes are typically chosen to go through all symmetrically nonequivalent bonds or sites in two mutually perpendicular directions. Any number of planes necessary for a reliable determination of the parameters can be defined. Although this is less automated than the generic schemes, it is convenient for studying features of special interest in the electric potential, such as lone pairs, hydrogen bonds, and polarization. By plotting equipotential lines on these planes, it is easy to compare the ab initio equipotential lines with those produced by the optimized parameters. This gives a better picture of the quality of the fit of the electric potential than the RMS deviation alone. We have found that parameters derived using the planes are better determined (smaller correlations and error limits) than those obtained using the geodesic point selection scheme.[25] Further, because the planes are defined using internal coordinates, they rotate with the molecule, and in contrast to most other methods, our procedure is therefore completely invariant to rotation of the molecule (or the Cartesian coordinate axes).

Polarizability parameters in our method are also determined by fitting to ab initio electric potentials. In this case external electric fields covering a suitable range of magnitude and direction are applied. This differs from approaches where polarizability parameters are determined from energy properties.[26] Because there are many effects that contribute to the interaction potential energy of a system, electrostatic and polarizability models and parameters can more safely be determined from ab initio electric potentials. The above approaches make for a consistent and robust

incorporation of polarizability into the description of the electrostatic component of V_{nb} in Eq. (6).

van der Waals Parameters. The optimization of vdW parameters is nontrivial. The primary reason for this is that the problem with correlations between the parameters is even worse than in the case of the electrostatic potential. To complicate things further, the optimization is now nonlinear and therefore requires initial values for the parameters. Also, because of the non-linearity, many different solutions are possible for the same set of parameters (multiple minima), and they can often give an equally good fit to the data. This has also been pointed out by others[27,28] in connection with optimization to observed data.

We are using systems such as dimers and complexes to optimize the vdW parameters. We reduce the correlation problems, in distinction to other similar methods,[26] by using QM intermolecular gradients as well as energies for the model systems, and by using a multitude of different intermolecular configurations and intermonomer distances. Single point ab initio calculations are set up with the dimers in carefully chosen configurations designed to effectively probe the vdW interactions of all the atoms for which parameters are to be determined. For each configuration, the intermonomer distance and orientation are varied over a range sufficient to yield good data on both the attractive and repulsive van der Waals interactions. Since we want the ab initio energy and gradient data to reflect intermolecular interactions only, we keep each monomer geometry fixed at the equilibrium structure of the isolated molecule. We account for the basis set superposition error in the dimers and complexes by using the counterpoise correction. When the ab initio data have been obtained, the vdW parameters are determined in a fit to all of the data simultaneously. In this procedure, the previously derived electrostatic parameters, and any already known vdW parameters, are held fixed. We test for uniqueness of the parameters, and/or best fit, by starting the optimization from different sets of initial values. We also check the optimized parameters for indeterminacy by calculating their statistical uncertainties and correlations using the covariance matrix. In order to obtain the vdW parameters for the peptide group, this procedure has been implemented for N-methylacetamide (NMA) dimers, producing parameters that are consistent with our optimized electrostatic parameters.[29]

As discussed above, now that V_{nb} is available, the valence force constants and intrinsic geometry parameters are readily determined from Eqs. (13) and (14), respectively. With all other parameters determined,

the torsion parameters can then be optimized to QM barrier profiles.

SDFF Advantages

There are a number of advantages to our procedurally consistent SDFF approach to refining an MM force field. By transforming a complete (i.e., ab initio) spectroscopic force field scaled to experimentally assigned bands, frequency agreement is assured at the start. In contrast to a least-squares fit, this not only directly incorporates highly accurate information (position and shape) about the minima in the potential surface, but it avoids the possibility of biasing parameters to compensate for an incomplete formulation of the valence terms in V. Correlations between valence parameters, a very serious problem in least-squares fitting, are also avoided. In addition, since the optimization can be done in a nonredundant coordinate basis,[20] the uniqueness of the force field is assured, and it is often possible to check that the MM interaction force constants are physically reasonable. Subsequent transformation to a redundant basis of choice is always possible. Also, there is a better chance of revealing the nature of specific conformation dependencies—for example, in the cross-terms—since no such dependence needs to be initially imposed on the MM force constants but can be deduced from the values obtained from many different conformers. Still another advantage is that nonphysical features in the nonbonded potential can be detected since they are likely to cause inconsistencies in the valence parameters obtained from the different molecules and conformations. Finally, accurate anharmonic contributions are obtained directly from the sampling of the varying bond lengths and angles in the different conformers. It is worth emphasizing that good reproduction of vibrational frequencies, besides being of importance for spectroscopy, is needed if one wants to compute reliable zero-point-corrected energies.

SDFF Implementation to Date

The SDFF protocol has been implemented for linear[9] and branched[13] saturated hydrocarbon chains, with excellent results. Ab initio geometries are reproduced within a few mÅ in bond lengths, within 1° in bond angles, and within 3° in torsion angles. Relative energies of conformers are reproduced with an average deviation of 0.2 kcal/mol, with discrepancies in barriers being <1 kcal/mol. For 791 non-CH stretch frequencies, the RMS deviation is ~ 6 cm^{-1} with respect to ab initio and ~ 5 cm^{-1} with respect to experimental frequencies. These SDFFs reproduced

the most reliable experimental elastic modulus of polyethylene to ~ 1%,[30] infrared intensities of alkanes significantly better than previous MM models,[31] and have enabled accurate zero-point-corrected energies to be calculated for the crystalline unit cells of forms I–IV syndiotactic polypropylene.[32] The SDFF results for olefinic hydrocarbon chains[11] are comparable in structures and energies, with an RMS deviation for 304 non-CH stretch frequencies of ~ 9 cm^{-1}.

ADVANCES IN THE SDFF REPRESENTATION

We have noted that insisting on spectroscopic accuracy as one of the constraints on an energy function can lead to a deeper understanding of the requirements on accepted and even new terms in the force field. In this section we briefly describe aspects of our recent developments in this area, particularly as they apply to peptide systems.

Torsions Around Unsaturated Bonds

Except for an early Lifson force field,[5] our previous CFFs,[33,34] and our current SDFF,[11] most force fields mistakenly use redundant torsion (t) and out-of-plane bend (ob) coordinates for unsaturated bonds, such as C=C bonds in alkenes and the CN bond of the peptide group. As a result, the torsion barrier and the torsion and out-of-plane bend frequencies cannot all be calculated correctly. For example, in the case of C=C the torsion barrier in ethylene is off by 14 kcal/mol[35] and the CH_2 twist frequency in isobutene is off by 63 cm^{-1} [36] (even though the MM4 force field[35] contains t/ob and ob/ob cross-terms). The problem is specifically due to the fact that, in the redundant representation, deformation of one of the out-of-plane coordinates also registers as a torsion. This leads to incorrect vibrational frequencies because the (nonredundant) torsion force constant is much larger than the out-of-plane bend force constant: for example, based on our SDFF,[11] we find that in ethylene the (nonredundant) C=C t force constant is more than twice as large as that of CH_2 ob.

These difficulties are eliminated by using Bell's torsion,[10] which, as mentioned above, is described as the sum or average of two of the dihedral angles that share the central bond but have different arms. This coordinate measures the angle between the π orbitals (in a direction perpendicular to the sp^2–sp^2 bond) and is orthogonal to the associated out-of-plane bend coordinates. With this coordinate implemented in our
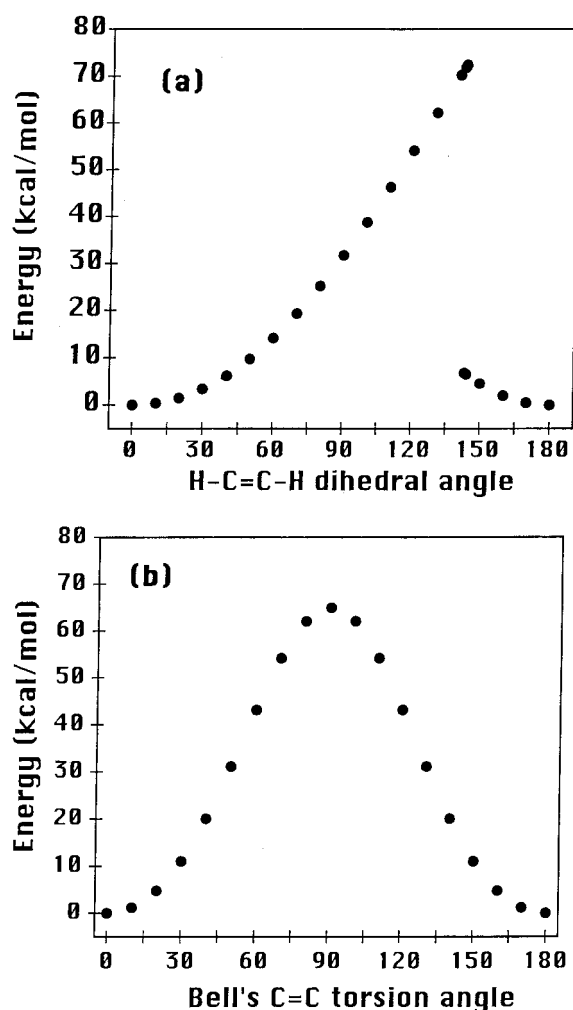

FIGURE 1 Ab initio [MCSCF-CAS(2,2)/6-31G(d)] torsion barrier in ethylene. Based on data from Ref. 11.[37] (a) Using redundant torsion coordinate (in degrees). (b) Using nonredundant torsion coordinate (in degrees).

alkene SDFF,[11] the deviation in the aforementioned CH_2 twist frequency of isobutene is only 5 cm^{-1} and in general all out-of-plane frequencies in the alkenes can be accurately reproduced using only the regular nearest-neighbor cross-terms. The torsion barrier profile, which cannot be computed simply by rotating one of the dihedral angles, is also correctly represented by using Bell's torsion. This is shown in Figure 1, with the redundant coordinate definition used in Figure 1a, and with our nonredundant procedure[11,37] used in Figure 1b.

It is interesting that the use of Bell's torsion leads to the illumination of other properties. Colthup and Orloff[38] noticed a "carbon electronegativity effect" in substituted ethylenes—viz., that the frequency of the CH_2 ob mode decreases as the electron density of the CH_2 carbon atom increases. Since this mode is almost

pure in each case, this means that the CH_2 ob force constant decreases with increasingly negative carbon charge. To see whether this is consistent with our alkene SDFF,[11] we calculated CHELPG charges for ethene, propene, and isobutene and obtained $-0.245e$, $-0.435e$, and $-0.582e$, respectively. This is in good agreement with our corresponding out-of-plane bend force constants of 36.6, 34.1, and 32.3 kcal/mol/rad^2. It is unlikely that such a correlation would be evident if redundant torsion and out-of-plane bend coordinates were used.

Similar considerations apply to the peptide group. For example, in Ref. 39, using redundant coordinates, the CN t frequency in NMA was off by 140 cm^{-1} and the "amino inversion" frequency was off by 22 cm^{-1} in NMA, by 100 cm^{-1} in N-methylformamide, and by 293 cm^{-1} in formamide. The difference in nonredundant force constants in NMA is even larger than in ethylene, the CN t force constant being more than nine times larger than the NH ob force constant. And the (ab initio) torsion barrier profile that results from the (redundant) rotation about the main chain CCNC angle of NMA (Figure 2a) is very different from the proper profile obtained using Bell's torsion coordinate (Figure 2b). The frequency discrepancies are also eliminated by this procedure (see below).

NH Out-of-Plane Bend

The use of nonredundant torsion and out-of-plane bend coordinates and the requirement of frequency accuracy has led us to a more detailed study of the interrelated behavior of these terms in the potential energy function.[40] In current force fields it is assumed that the NH ob coordinate is *intrinsically* planar, so its reference value is set to zero. This is undoubtedly correct as long as the CN t coordinate is also planar, with the peptide group resonance at its maximum. However, our results[40] indicate that, for nonplanar values of the CN t coordinate, the reference value of the NH ob angle should not be zero but should vary with the torsion angle. The ab initio [MP2/6-31++G(d,p)] NH ob potential energy in NMA for two different values of CN t is shown in Figure 3. This dependence will not only be important in obtaining accurate frequencies but can be expected to influence molecular dynamics simulations, which in turn, through the dynamics of the NH bond vector, could be important in the interpretation of NMR spectra.[41]

In connection with these studies, we noticed that there are large charge fluxes associated with the NH ob deformation in NMA,[40] the largest charge variation occurring on the N atom. This effect, which is $\sim 25\%$ when the out-of-plane angle varies between 0°
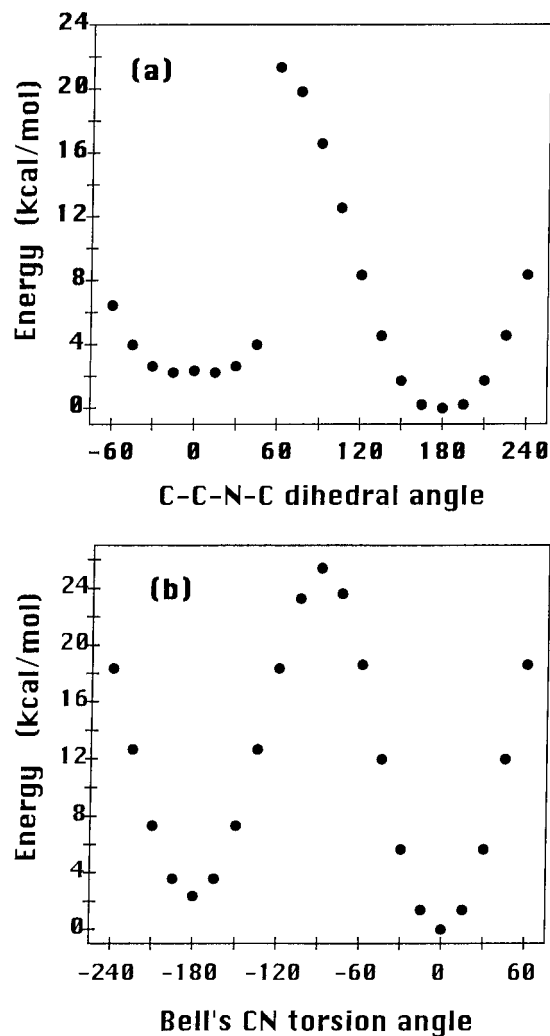

FIGURE 2 Ab initio [MP2/6-31++G(d,p)] torsion barrier about peptide bond in N-methylacetamide. (a) Using redundant torsion coordinate (in degrees). (b) Using nonredundant torsion coordinate (in degrees).

and 45°, is shown in Figure 4. Since such deformation requires relatively little energy, this charge variation should be included in simulations if they are to be realistic, particularly in view of the importance of the peptide group electrostatic interactions in determining the structures of proteins. The effect is not an artifact of the particular method used to optimize potential derived charges. This is clearly shown by the fact that very similar results are obtained with three different methods: CHELPG,[42] the geodesic point selection scheme,[25] and our "planes through the molecule" method.[14,29] Nor can the effect be explained by polarization through the electric field, which we have implemented.[14] The only way to account for the charge variation is to have the charges depend explicitly on the out-of-plane bend coordinate.
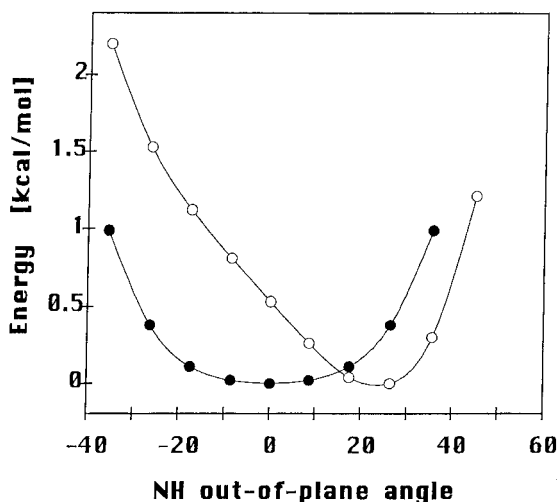
FIGURE 3 Ab initio potential energy for NH out-of-plane bend deformation in N-methylacetamide as a function of CN torsion angle. (●) CN torsion angle = 0°. (○) CN torsion angle = 15° (higher energy minimum offset by 1.14 kcal/mol).

SDFF for the Peptide Group

Pursuant to our goal of producing an SDFF for the polypeptide chain, we have derived an intermediate SDFF for the peptide group of NMA. This force field uses our published fixed atomic point charge model,[29] the recently derived vdW parameters for NMA,[43] and the SDFF-transformed HF/6-31+G(d) scaled[44] force constants, geometry, and dipole derivatives. In Table I the MM vibrational frequencies are compared with the (scaled) ab initio and observed[45–47] results. The SDFF valence parameters were obtained directly by the analytical transformation. For each CH_3 group, the parameters that are topologically identical (e.g., the three NCH bend force constants) were averaged, and the peptide CN and the CH_3 cosine torsion potentials were constructed to be in agreement with the torsion force constants and equilibrium geometries given by the transformation. No least-squares fitting of valence parameters to frequencies or geometries was done. Charge flux and internal coordinate dipole flux parameters compatible with our previous model for calculating ir intensities[31] were obtained by optimization to the ab initio dipole derivatives (although they were not included in the energy function). The SDFF and ab initio intensities are compared in Figure 5, in which the bands are Lorentzians plotted with a 10 cm^{-1} half-width for all lines.

The results are very satisfactory. The ab initio geometry is reproduced to RMS errors of 0.9 mÅ in bond lengths, 0.5° in bond angles, 1.0° in torsion angles, and 0.1° in out-of-plane angles. The frequency

agreement is equally good, with an RMS error of 8.5 cm^{-1}. The agreement would be much better if not for the large discrepancies in the modes at 1120 (28 cm^{-1}) and 1264 (20 cm^{-1}), which are due to the present averaging of force constants associated with each CH_3 group. This is seen from the fact that the errors disappear if the exact truncated force field is used, i.e., a force field in which all the included force constants are allowed to retain their initial SDFF-transformed values. The inclusion of charge fluxes in the frequency calculation may cure this problem, or it may be necessary to implement explicit conformation dependence for the affected force constants. Finally, the reproduction of ir intensities is excellent, showing that the SDFF provides a faithful reproduction of eigenvectors.

CONCLUSIONS

As we have observed, Lifson made an effort to "combine conformational analysis and vibrational analysis into a united self-consistent procedure,"[2] a goal that has not been sought by most subsequent developers of energy functions and their parameters. A reason often given for avoiding this approach is that the (mainly "higher") frequencies are not crucial for understanding the important low-frequency motions that govern macromolecular properties in general and biological properties in particular. Lifson's philosophy, and one that we promote in our SDFF procedure, is that, for many reasons, this point of view is not fruitful in leading to the development of a physically reliable
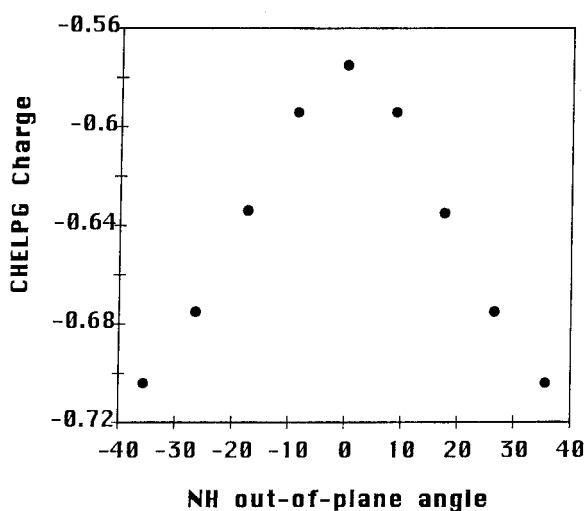

FIGURE 4 CHELPG charge (in electrons) on N atom of N-methylacetamide as a function of NH out-of-plane angle (in degrees) at CN torsion angle = 0°.

Table I   Observed, SDFF, and Scaled Ab Initio [HF/6-31+G(d)] Frequencies of N-Methylacetamide (in cm^{-1})

ν (Obs.)[a]	ν (SDFF)	ν (Ab Initio)	$\Delta\nu$[b]	Sym.[c]	Potential Energy Distribution[d]
3507[e]	3507	3507	0	A′	NH s(100)
3008	3001	3002	−1	A′	NCH$_3$ as(100)
3008	3001	2995	6	A″	CCH$_3$ as(100)
2973	2994	2992	2	A″	NCH$_3$ as(100)
2973	2990	2989	1	A′	CCH$_3$ as(99)
2958	2929	2931	−2	A′	NCH$_3$ ss(100)
2915	2928	2925	3	A′	CCH$_3$ ss(99)
1706	1707	1708	−1	A′	CO s(83) CCN d(11)
1511	1513	1512	1	A′	NH ib(51) CN s(28)
1472	1459	1469	−10	A′	NCH$_3$ ab(80)
1446	1437	1445	−8	A′	CCH$_3$ ab(69) CCH$_3$ sb(11)
1446	1440	1442	−2	A″	NCH$_3$ ab(94)
1432	1422	1433	−11	A″	CCH$_3$ ab(92)
1419	1418	1423	−5	A′	NCH$_3$ sb(96)
1370	1372	1379	−7	A′	CCH$_3$ sb(79) CCH$_3$ ab(15)
1265	1284	1264	20	A′	CO ib(21) NH ib(21) CN s(18) CC s(10) CCH$_3$ sb(10)
1181	1181	1176	5	A′	NCH$_3$ r(40) NC s(12)
	1092	1120	−28	A″	NCH$_3$ r(89)
1089	1101	1094	7	A′	NC s(56) NCH$_3$ r(12)
1037	1036	1040	−4	A″	CCH$_3$ r(63) CO ob(19)
990	994	985	9	A′	CCH$_3$ r(52) CC s(21)
857	862	851	11	A′	CN s(35) NCH$_3$ r(20) CCN d(11)
658	647	648	−1	A′	CC s(36) CO ib(34)
626	622	627	−5	A″	CO ob(67) CCH$_3$ r(26) CN t(11)
429	449	449	0	A′	CCN d(51) CO ib(28) CCH$_3$ r(14)
391[e]	394	391	3	A″	CN t(118) NH ob(51) CO ob(23)
279[e]	279	279	0	A′	CNC d(67) CCN d(28)
	149	162	−13	A″	NH ob(65)
	74	70	4	A″	NC t(254) NH ob(116)
	32	28	4	A″	CC t(400) CO ob(155)

[a] Reference 45.

[b] $\Delta\nu = \nu$ (SDFF) $- \nu$ (ab initio).

[c] Symmetry species.

[d] Potential energy distributions (contributions $\geq$ 10). Symmetry coordinates follow our previous definition.[48] s: Stretch, ss: symmetric stretch, as: antisymmetric stretch, sb: symmetric bend, ab: antisymmetric bend, d: deformation, r: rock, t: torsion, ib: in-plane bend, ob: out-of-plane bend.

[e] References 46 and 47.

energy function, i.e., one that is "an empirical representation of the Born-Oppenheimer approximation."[2] In fact, the goal of the SDFF methodology is to reproduce this approximation in its exact sense, as represented by ab initio-calculated molecular "data," leaving for molecular dynamics (or Monte Carlo) simulations the calculation of most condensed phase properties.

As developed in its theoretical formulation,[7,8] and reviewed briefly in this paper, the SDFF method achieves this result by requiring consistency in the parameter set and frequency agreement as an initial constraint, together with structure and energy agreement. In the optimization of the potential energy func-

tion (form and parameters), this is accomplished by an exact analytical transformation of ab initio "data" into the molecular mechanics format. We have shown that this protocol also provides a very sensitive means to elucidate basic physical components of the energy function, and our studies have included in this respect the use of Bell's torsion for unsaturated bonds,[11] the representation of polarization,[14,29] the necessity of incorporating charge fluxes,[31,40] the existence of a nonplanar intrinsic NH out-of-plane bend angle for nonplanar peptide CN torsion angles,[40] and a new description of hydrogen-bond formation.[12]

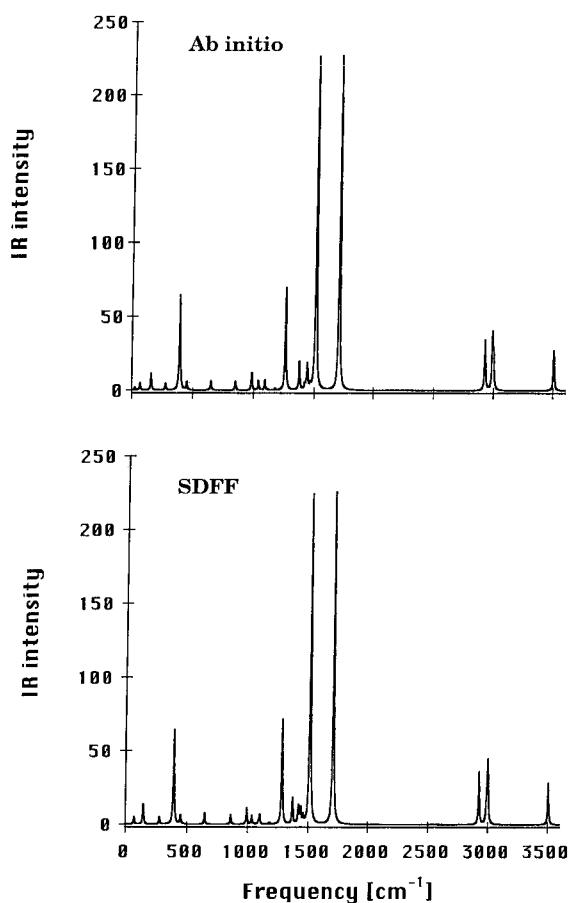The resulting SDFFs obtained thus far for saturated and unsaturated hydrocarbons[9,11,13] give excellent re-

FIGURE 5 Calculated ir spectra of N-methylacetamide by ab initio and based on our SDFF force field with charge and dipole flux parameters.

production of the ab initio structures, energies, and vibrations: bond lengths to a few mÅ, bond angles to ~ 1°, torsion angles to ~ 3°, relative energies to a few tenths of a kcal/mol, energy barriers to <1 kcal/mol, and frequencies to RMS deviations (for non-CH stretch modes) of 5–10 cm^{-1}. Preliminary to a full SDFF for the polypeptide chain, the results for NMA are comparable, with an RMS frequency deviation of 8.5 cm^{-1} and excellent reproduction of the ab initio ir intensities.

The SDFF methodology is an extension and elaboration of the Lifson vision of energy function development and parameter optimization, and provides a consistent and robust route to a classical representation of a quantum-mechanical potential energy surface.

Note Added in Proof: In connection with the summing of nonbonded interactions over all atom pairs in positions 1,4 and higher (see Eq. 6), we noted that it is important that there be no unbalanced charges in the

summation of Coulomb interactions, since this results in unphysical electrostatic energies. We have developed a simple method for avoiding this problem (K. Palmo, B. Mannfors, and S. Krimm, Chem. Phys. Lett., submitted), which also leads to more reasonable and transferable torsion potentials.

REFERENCES

1. Baker, D.; Sali, A. Science 2001, 294, 93–96.
2. Lifson, S.; Warshel, A. J Chem Phys 1968, 49, 5116–5129.
3. Warshel, A.; Lifson, S. J Chem Phys 1970, 53, 582–594.
4. Warshel, A.; Levitt, M.; Lifson, S. J Mol Spectrosc 1970, 33, 84–99.
5. Ermer, O.; Lifson, S. J Am Chem Soc 1973, 95, 4121–4132.
6. Lifson, S.; Stern, P. S. J Chem Phys 1982, 77, 4542–4550.
7. Palmo, K.; Pietila, L.-O.; Krimm, S. Comput Chem 1991, 15, 249–250.
8. Palmo, K.; Pietila, L.-O.; Krimm, S. J Comput Chem 1991, 12, 385–390.
9. Palmo, K.; Mirkin, N. G.; Pietila, L.-O.; Krimm, S. Macromolecules 1993, 26, 6831–6840.
10. Bell, B. P. Trans Faraday Soc 1945, 41, 293–295.
11. Mannfors, B.; Sundius, T.; Palmo, K.; Pietila, L.-O.; Krimm, S. J Mol Struct 2000, 521, 49–75.
12. Qian, W.; Krimm, S. J Phys Chem A 2002, 106, 6628–6636.
13. Palmo, K.; Mirkin, N. G.; Krimm, S. J Phys Chem A 1998, 102, 6448–6456.
14. Mannfors, B.; Palmo, K.; Krimm, S. J Mol Struct 2000, 556, 1–21.
15. Ewig, C. S.; Waldman, M.; Maple, J. R. J Phys Chem A 2002, 106, 326–334.
16. Van Duijnen, P. Th.; Swart, M. J Phys Chem A 1998, 102, 2399–2407.
17. Palmo, K.; Pietila, L.-O.; Krimm, S. Computers Chem 1993, 17, 67–72.
18. Mirkin, N. G.; Krimm, S. J Phys Chem 1993, 97, 13887–13895.
19. Palmo, K.; Krimm, S. SPEAR: Spectroscopic Potential Energy Analytical Refinement—A Molecular Parametrization and Modeling Package. To be published.
20. Palmo, K.; Pietila, L.-O.; Krimm, S. J Comput Chem 1992, 13, 1142–1150.
21. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269–10280.
22. Sigfridsson, E.; Ryde, U. J Comput Chem 1998, 19, 377–395.
23. Bush, B. L.; Bayly, C. I.; Halgren, T. A. J Comput Chem 1999, 20, 1495–1516.

24. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. J Comput Chem 1993, 14, 1347–1363.

25. Spackman, M. A. J Comput Chem 1996, 17, 1–18.

26. Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. J Phys Chem A 1999, 103, 9872–9882.

27. Dinur, U.; Hagler, A. T. J Am Chem Soc 1989, 111, 5149–5151.

28. Rizzo, R.; Jorgensen, W. L. J Am Chem Soc 1999, 121, 4827–4836.

29. Mannfors, B.; Mirkin, N. G.; Palmo, K.; Krimm, S. J Comput Chem 2001, 22, 1933–1943.

30. Palmo, K.; Krimm, S. J Polym Sci Polym Phys Ed. 1996, 34, 1–9.

31. Palmo, K.; Krimm, S. J Comput Chem 1998, 19, 754–768.

32. Palmo, K.; Krimm, S. Macromolecules 2002, 35, 394–402.

33. Mannfors, B.; Pietila, L.-O.; Palmo, K. J Mol Struct 1991, 248, 289–316.

34. Mannfors, B.; Pietila, L.-O.; Palmo, K. J Mol Struct 1994, 328, 287–295.

35. Nevins, N.; Chen, K.; Allinger, N. L. J Comput Chem 1996, 17, 669–694.

36. Nevins, N.; Allinger, N. L. J Comput Chem 1996, 17, 730–746.

37. Sundius, T. Personal communication.

38. Colthup, N. B.; Orloff, M. K. Spectrochim Acta 1971, 27A, 1299–1314.

39. Maple, J. R.; Hwang, M. J.; Jalkanen, K. J.; Stockfish, T. P.; Hagler, A. T. J Comput Chem 1998, 19, 430–458 (supplementary material).

40. Mannfors, B.; Mirkin, N. G.; Palmo, K.; Krimm, S. J Phys Chem A, submitted.

41. Buck, M.; Karplus, M. J Am Chem Soc 1999, 121, 9645–9658.

42. Breneman, C. M.; Wiberg, K. B. J Comput Chem 1990, 11, 361–373.

43. Mirkin, N. G.; Palmo, K.; Krimm, S. To be published.

44. Mirkin, N. G. Unpublished.

45. Ataka, S.; Takeuchi, H.; Tasumi, M. J Mol Struct 1984, 113, 147–160.

46. Fillaux, F.; Lozé, C. J Chim Phys 1976, 73, 1004–1009.

47. Fillaux, F.; Lozé, C. J Chim Phys 1976, 73, 1010–1017.

48. Mirkin, N. G.; Krimm, S. J Mol Struct 1991, 242, 143–160.