On the Transferability of Force Field Parameters—With an ab Initio Force Field Developed for Sulfonamides

F. Sato and S. Hojo

Ryoka Systems Inc., Urayasu, Chiba, Japan

H. Sun*

Aeon Technology Inc., San Diego, California 92129

Received: July 25, 2002; In Final Form: October 30, 2002

With 25 organic molecules that represent popular functional groups, we tested the transferability of force field parameters using both parametrized force fields CHARMM, CFF, and MMFF and generic force fields DREIDING and UNIVERSAL. We found that, if transferred parameters are used in a parametrized force field, the calculation quality is no longer superior to that of a generic force field. To achieve high quality in predictions, new parameters should be created from ab initio data whenever necessary. We investigated this approach and found that a custom-built force field can be made to reproduce ab initio results if parameters are derived specifically for the molecules of interest. The parametrization procedure was applied to a group of classic antibacterial drug molecules, the sulfonamides.

I. Introduction

Force field applications have been extensively based on transferability of force field parameters, which assumes the parameters derived from a training set of model molecules can be applied to molecules with similar chemical structures.[1−3] Although this assumption is generally valid based on observations that many molecular properties are approximately transferable among similar molecular environments, how the force field parameters are transferred and what the consequences of the transfer should be expected to need clarification.

Generally speaking, there are two types of parameter transfers. One is internal, in which the force field parameters are transferred within a molecule. For example, parameters derived from monomers or residues can be used for macromolecules.[4−6] This type of transfer is valid in most cases. The other is external, where parameters derived from one molecule are used for similar but different molecules. For example, parameters derived for alkanes might be used for halogen-substituted alkanes. The external transfer could introduce considerable errors simply because some of the molecular properties (e.g., structural parameters) may not be strictly transferable when some of the atoms are substituted.

The problems are largely due to the use of atom types underlying force field methods. An atom type is an alphanumerical string that represents a "class" of atoms in a certain environment.[1−6] Each element may be classified into several atom types in a force field, depending on its chemical environment. Force field parameters for a system are identified completely based on its atom types. Correct usage of atom types is thus a key factor in the success of force field methods. However, an atom type is usually defined by looking at atomic attributes including element symbol, connectivity, hybridization, etc., and sometimes the attributes of the nearest neighbor atoms. Consequently, they are generally defined for the very small, local environment presented in a training set for parametrization. When they are applied to molecules outside of that training set, atoms in environments differing in ways not accounted for by the training set could be assigned the same atom type, which then leads to the use of the same force field parameters and, subsequently, errors in calculation. The force field thus could not be reliably applied to molecules outside of its training set, which by definition breaks its transferability.

Unfortunately, how force field parameters are transferred depends on software implementation. Usually, parameters are transferred externally with few or no checks on the validity by software packages that are popularly used today. The users must often use their own judgment to decide whether the calculated force field parameters are valid and usually only after the simulations are done. We believe this causes the majority of serious errors and uncertainties in force field applications, which hinders the broad use of force field technology in drug discoveries and material designs. To our best knowledge, no published work to date specifically addresses this issue. Almost all published force field work focuses on molecules that are either used as training sets or very closely resemble the training set. This motivated us to carry out this study. We tested the transferability by applying several common force fields to molecules that were unlikely to have been used as the training sets in these force field developments. The first question we attempted to address is how large would the errors be quantitatively when parameters are transferred?

It became clear, as will be discussed in this paper, significant errors could be obtained when parameters are transferred externally. Because the variation of chemistry is so large that it is unlikely a single force field could be made to meet every need, custom-built force fields would be necessary whenever the required parameters are missing. The second objective of

* To whom correspondence should be addressed.

10.1021/jp026612i CCC: $25.00
Published on Web 12/17/2002

this study is to seek a solution when parameters of interest are found to be nontransferable.

Methods of deriving force field parameters from ab initio data were established in the late 80s. Among the many publications in this field, Hagler et al.[7−9] have a series of papers on CFF force field development. Halgren[10] applied this technique to make the MMFF force field. Dasgupta et al.[11] published their work using a similar approach. However, to construct a force field from ab initio data is not a trivial task. The challenge arises from several factors associated with the least-squares fitting process: ambiguous combinations of parameters defined by redundant internal coordinates, employment of inadequate functional forms to represent the energy surfaces, and incomplete sampling of data points on the surfaces. Generally speaking, the fit is both over-determined (more data points than variable parameters) and under-determined (ambiguous combination of correlated parameters), and the ambiguities are often unknown prior to the fit process. Simply applying the least-squares fit usually does not work. Expert developers often apply various empirical controls to guide the fit, which makes the procedure very tedious and, in most cases, very difficult for a novice developer.

A new software product, *Direct Force Field*, has been developed recently.[12] Unlike the previous approach, this product is implemented with procedures that an expert developer would use to manually fit a force field. This tool could be used to generate force field parameters rapidly from ab initio data. In this work, we utilized this tool to generate new force field parameters for a set of testing molecules to check if a rapid development of force field parameters is feasible.

Finally, to investigate how the new software and techniques can be utilized in real life applications, we parametrized a group of antibacterial drug molecules: sulfonamides. The sulfonamide family is important because it represents the first real success in treating bacterial infections with relatively safe, nontoxic chemical compounds. Discovered in the 1930s, these drugs are still in use and widely referenced in textbooks in explaining concepts of competitive enzyme inhibition and antimetabolites,[13] which are used to guide the discovery of new drug molecules.[14]

There have been several previous force field publications regarding this system.[15,16] However, published parametrization work is limited to small fragments of sulfonamides, which do not provide enough parameters for real drug molecules. A high quality force field designed for this family of molecules would be a valuable contribution to continued research in this field.

II. Methods

2.1. Ab Initio Calculations. In this project, ab initio calculations were used for generating reference data when experimental data was not available for comparison, and for generating data from which force field parameters were derived. All ab initio calculations were conducted using the software package *Gaussian 98*.[17] The following combinations of methods and basis functions were used for different purposes: Hartree−Fock method[18] with the 6-31G(d) basis set,[19] Møller−Plesset second-order perturbation method (MP2)[20] with the 6-311G-(d,p) basis sets,[21] and Becke's three parameter hybrid functional using the LYP correlation functional (B3LYP) method[22] with the 6-311G(d,p) basis set.

Full geometry optimizations using analytical gradients were performed to characterize minimum-energy structures. The optimized structures were verified using the normal-mode analysis. Total energies, analytical gradients, and Hessian matrixes (first and second derivatives of the total energies) were calculated in order to prepare data for parametrization. In addition to structures with minimized energies, distorted structures in which the dihedral angle of interest was fixed at selected values to sample conformational spaces were also calculated. Atomic partial charges were calculated from wave functions using Mulliken population analysis and by fitting to electrostatic potentials using the Merz−Singh−Kollman scheme.[23,24]

2.2. Molecular Mechanics and Dynamics. Molecular mechanics energy minimization and dynamics simulations were carried out using built-in force fields from software packages *Cerius*[2] [25] and *MOE*.[26] In testing force field transferability, calculations were conducted with atom types assigned using default procedures provided in the software packages.

Molecular dynamics (MD) simulations were carried out using the software package *Discovery*.[27] For crystals, a super cell consisting of several unit cells with explicit minimum image convention was used. The cell edges ranged from 20 to 30 Å in length and contained 1400−3600 atoms. The Ewald summation method[28] was used for both van der Waals (VDW) and electrostatic terms. Constant pressure and temperature (NPT) simulations were carried out using a modified velocity Verlet integrator[29] with Andersen temperature and pressure control method[30] for validation calculations. Conditions used in all MD simulations were the time step of 1.0 fs, the temperature of 298 K, and the pressure of 0.100 MPa. The equilibration took about 18 ps, which is usually adequate for crystals. The average periods were 2 ps for the NPT simulations.

2.3. Force Field Parametrization. Force field parametrizations were conducted using the software package, *Direct Force Field*,[12] which derives force field parameters from ab initio data by minimizing the chi-square quantity (least-squares methods):

$$\chi^2 = \sum_{i=1}^{N}\left[\frac{y_i - y(x_i;a_1,a_2,\cdots,a_M)}{\sigma_i}\right]^2 \qquad (1)$$

The input data y_i includes atomic charges, energies, and the first and second derivatives of the energies. The parameters a_k are force field parameters to be adjusted, which are usually classified into valence, charge and VDW parameters. The "weighting factors" σ_i normalize the quantities. Normally, the number of data points (N) is much greater than the number of variables (M).

The functional forms $y(x_i;a_1,a_2,\cdots,a_M)$ we used in this work include a set of CHARMM type of functions:[4]

$$E = \sum_{\text{bond}} K_b(b - b_o)^2 + \sum_{\text{angle}} K_\theta(\theta - \theta_o)^2 + \sum_{\text{UB}} K_{\text{UB}}(S - S_o)^2 + \sum_{\text{dihedral}} K_\phi[1 + \cos(n\phi - \phi_o)] + \sum_{\text{improper}} K_\chi(\chi - \chi_o)^2 \sum_{\text{nonbond}}\left\{\epsilon_{ij}\left[\left(\frac{R_{ij}^o}{R_{ij}}\right)^{12} - 2\left(\frac{R_{ij}^o}{R_{ij}}\right)^6\right] + \frac{q_i q_j}{R_{ij}}\right\} \qquad (2)$$

with arithmetic combination rules for the VDW parameters:

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$
$$R_{ij}^o = \frac{R_i^o + R_j^o}{2} \qquad (3)$$

and simplified CFF type forms:[7−9]

$$E_{\text{total}} = \sum_{\text{bond}} \lfloor k_2(b - b_o)^2 + k_3(b - b_o)^3 + k_4(b - b_o)^4 \rfloor + \sum_{\text{angle}} [k_2(\theta - \theta_o)^2 + k_3(\theta - \theta_o)^3 + k_4(\theta - \theta_o)^4] + \sum_{\text{torsion}} [k_1(1 - \cos\phi) + k_2(1 - \cos 2\phi) + k_3(1 - \cos 3\phi)] + \sum_{\text{out-of-plane}} k_2(\chi - \chi_o)^2 + \sum_{\text{bond-bond}} k(b - b_o)(b' - b'_o) + \sum_{\text{bond-angle}} k(b - b_o)(\theta - \theta_o) + \sum_{i<j} \epsilon_{ij}\left[2\left(\frac{r_{ij}^o}{r_{ij}}\right)^9 - 3\left(\frac{r_{ij}^o}{r_{ij}}\right)^6\right] + \sum_{i<j} \frac{q_i q_j}{r_{ij}} \quad (4)$$

with 6-order combination rules:[8]

$$r_{i,j}^o = \left(\frac{(r_i^o) + (r_j^o)^6}{2}\right)^{1/6}$$

$$\epsilon_{i,j} = 2\sqrt{\epsilon_i \epsilon_j}\left(\frac{(r_i^o)^3 (r_j^o)^3}{(r_i^o)^6 + (r_j^o)^6}\right) \quad (5)$$

Internal coordinates used are bond lengths (b), bond angles (θ), Urey−Bradley 1−3 distance (S), torsion dihedral angles (ϕ), and out-of-plane angles or improper dihedral angles (χ). The cross-coupling terms used in CFF type force fields are known as being important for predicting vibration frequencies and structural variations associated with conformational changes. The nonbond interactions, which include the VDW terms and electrostatic interaction terms, are used for interactions between pairs of atoms that are separated by two or more intervening atoms or those that belong to different molecules. The electrostatic interactions are represented using atomic partial charges.

Although VDW parameters could be derived from ab initio data, it is accepted that a reliable and accurate way to derive those parameters is to use condensed phase simulations.[6,29] Therefore, we initially fixed VDW parameters by default values taken from literature and fitted only valence and charge parameters. After valence and charge parameters were optimized, VDW parameters were subject to optimization using MD simulations of condensed phases. The optimization procedure was repeated until a consistent fit was obtained for both gaseous and condensed phases.[6]

The χ^2 in eq 1 only measures how closely data fits but does not provide complete information on the performance of result parameters. A more sensible test of resulting parameters is to use these parameters to calculate certain molecular properties. These calculated properties can then be compared to experimental data and/or properties calculated using higher-level ab initio methods. The properties we calculated for validation purposes are usually molecular structures, conformational energies and structures, vibrational frequencies, and condensed phase properties such as density and cohesive energies.

III. Results and Discussion

3.1. Testing Transferability. A total of 25 organic molecules representing popular functional groups as illustrated in Figure 1 were used for testing parameter transferability. These compounds can be grouped into seven categories. Compounds 1−4 are hydrocarbons; 5−7 are benzene derivatives; 8−11 are carbonyl compounds; 12−17 are aromatic heterocyclic compounds; 18−20 are nonaromatic heterocyclics; the sulfon compounds 21−22 are included for the purpose of the parametrization work presented in this paper; and finally, compounds 23−25 are common ionic organic functional groups.

The testing molecules were selected based on the following criteria: (1) They represent popular functional groups, yet they are not basic compounds that are likely to have been used in training sets for parametrizations. In other words, some of the parameters for these molecules must be transferred. (2) High quality gas-phase experimental data are available for most of these compounds (molecules 1−20)[31] so that accurate comparisons can be made. In the cases where gas-phase experimental data is not available (for molecules 21−25), we supplemented the data using MP2/6-311G(d,p) data. In this paper, we refer to all baseline data as "reference data."

Five popular force fields were selected for testing: CHARMM,[5] CFF91,[8] MMFF(94),[10] DREIDING(2.21),[32] and UNIVERSAL (1.02).[33] The former three are parametrized; the latter two are generic. It should be emphasized that the purpose of this work is not to validate any of these force fields. Rather, we selected these force fields because they are generally available and represent differing coverage of molecules. CHARMM is mostly parametrized for amino acids and nucleic acids, CFF91 is well suited to common organics and proteins, and MMFF force field is perhaps the most extensively parametrized force field for organic molecules among the three parametrized force fields selected. DREIDING is a generic force field that is made empirically, rather than being specifically parametrized. Finally, UNIVERSAL represents a completely different approach; parameters are made based on a set of rules from atomic properties. It is very generic, covers any molecule consisting of any atoms from the periodic table.

We optimized the structures of test molecules using the selected force fields and compared calculated bond lengths and angles against the reference data. The comparisons are plotted in Figure 2, parts a and b, in which the reference data occupies the x coordinate and the calculated results are plotted on the y coordinates. A point falling on the diagonal line indicates a perfect match between the calculated and reference data. Different shapes of the dots represent results obtained using different force fields. As shown in these figures, although there is a clear correlation between calculated and reference data, the dots are widely scattered on both sides of the diagonal line.

A summary of percentage errors in bond lengths and angles calculated using different force fields are given in Table 1. Note that significant errors are found across the table for both bond lengths and angles using any of the force fields. Despite some fluctuations, which are related to the samples used in the study, all five force fields perform similarly. The percentage errors are as large as 18% in bond lengths and close to 13% in bond angles. More importantly, the parametrized force fields (CHARMM, CFF, and MMFF) are not significantly better than the generic force fields (UFF and DREIDING) based on the data calculated for the testing molecules.

Maximum percentage errors in both bond lengths and angles obtained for each of the molecules using different force fields are given in Table 2. Symmetries (point group symbol) in equilibrium structure for each of the molecules are also given in the table. It is of interest to point out that for some of these molecules one or two force fields can perform a very good prediction; the largest percentage errors can be as low as 1% in some of these cases. However, errors are scattered randomly in this table; there is no clear pattern, especially the three parametrized ones (CHARMM, CFF, and MMFF) of interest,

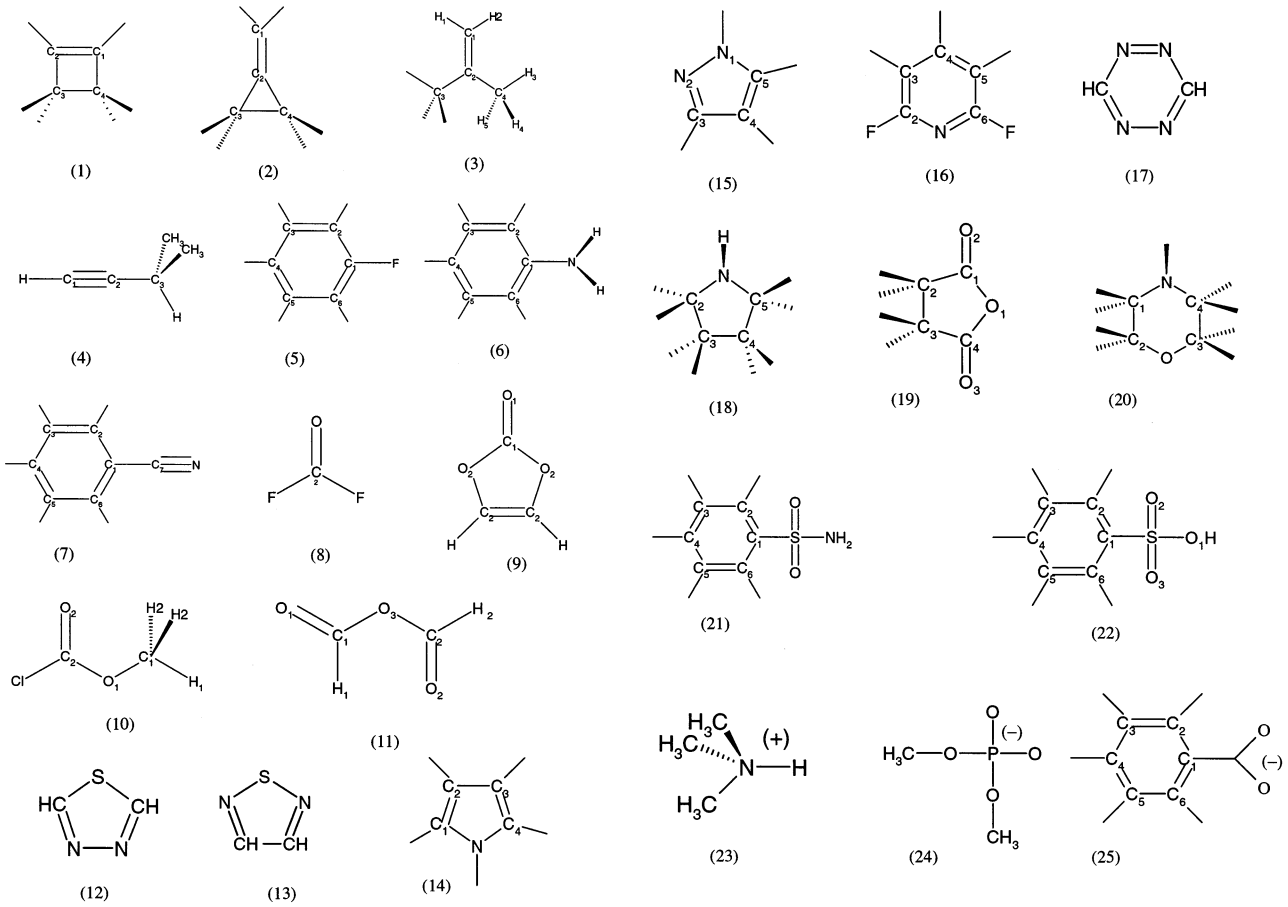

Figure 1. Model compounds selected for testing force field transferability.

showing either a group of molecules that can be treated in a similar precision by all force fields or a force field that can handle all of these molecules significantly better than others. That is, different force fields perform differently on the molecules tested.

It should be emphasized that the statistical data presented in Tables 1 and 2 do not truly represent the quality of the force fields tested because results depend on sample selection. It is well established that the CFF, CHARMM, and MMFF force fields have been rigorously validated for many molecules. The point we are addressing here is, rather, that a force field only works for what it is designed for. Once a force field is used for molecules that are not specifically parametrized, the quality of prediction falls into the same level of a generic force field. It is dangerous to assume a force field validated for other (even similar) molecules can be used without validation and to expect the same quality of prediction in these unvalidated molecules. Transferability cannot be assumed.

Only bond lengths and angles are used for the above discussions. This is because these properties are the most basic ones, which must be validated before probing other properties such as conformational energies, vibrational frequencies, and condensed phase properties. We found some evidences showing that errors in conformational energies and structures could be even greater because some optimized structures show broken symmetries. If conformational and vibrational properties were included, the disagreements between the force field and reference data would be more dramatic than what we obtained for the bond lengths and angles. Therefore, our conclusions remain unaltered.

3.2. Rapid Parametrization from ab Initio Data. Naturally, a solution to the problem of poor transferability is to parametrize a new force field whenever it is necessary. If parametrization can be done rapidly and reliably, each molecule (or molecular fragment) can be modeled using a set of unique parameters without transferring them externally.

Our first step is to investigate if simple force fields that can accurately predict structural properties can be made from ab initio data for the 25 test molecules. We simplified the parametrization issue by using only optimized structures without considering conformational variations and performed calculations using a modest ab initio calculation method, HF/6-31G-(d). The focus here is to determine how accurate a force field could be without transferring or sharing any parameters externally, provided the intrinsic limitations (approximations of atom types and functional forms, etc.) of the force field method.

Using the HF/6-31G(d) method, we optimized the 25 test molecules. We then calculated the ESP atomic charges, energies, gradients, and Hessian matrixes of the optimized structures and used them as input data in force field parameter derivations.

The atom types are defined generically using the following rules:

(1) The first one or two characters are the element symbol;

(2) It is followed by a character which is an integer (0−9) indicating coordination number (i.e., how many bonds are attached to this atom);

(3) Then a character indicates special situations that the atom is in, for example: “c”, in a small cyclic molecule; “r”, in resonant structure; “a”, in an aromatic ring; “p”, in a highly polarized environment (e.g., high charge nearby); “+”, for cations and “-” for anions.

The VDW parameters were fixed to a set of default values during the fit. The functional forms used in this study are CHARMM functional forms (eq 2 and 3). The weighting factors

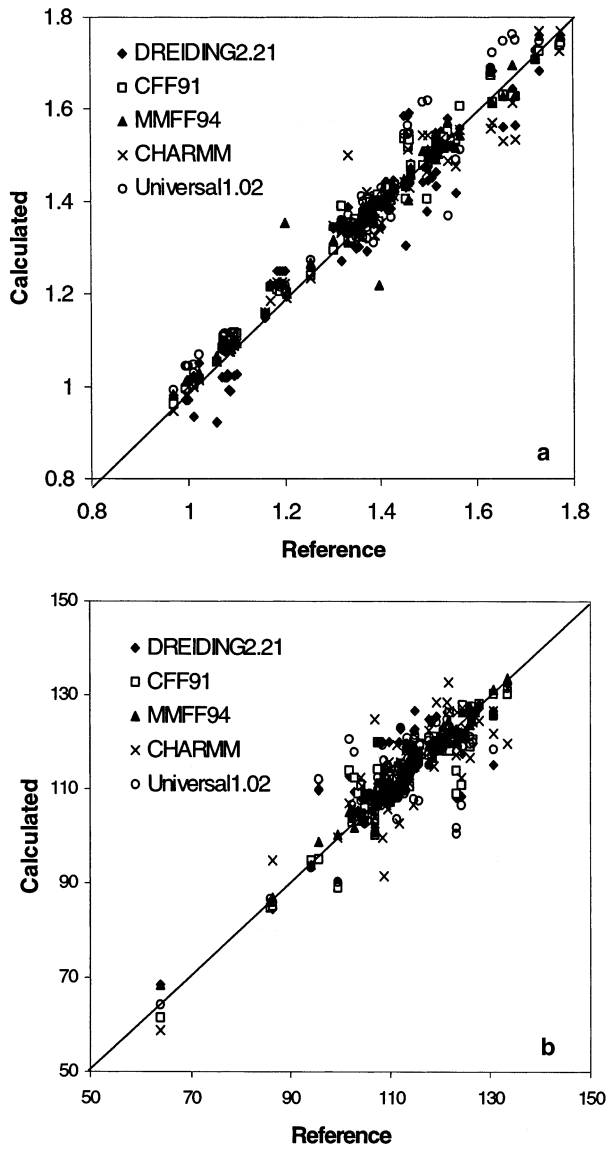

Figure 2. Comparison of calculated and reference bond lengths (a) and angles (b). The calculated data are obtained using different force fields. The bond lengths are in angstrom, and the angles are in degrees.

TABLE 1: Summary of Percentage Errors in Calculated Bond Lengths and Angles Using Different Force Fields[a] for the Testing Molecules

	bond lengths			bond angles		
force field	max(+)	max(−)	STD	max(+)	max(−)	STD
D	14.7	−12.5	4.5	9.6	−12.9	3.9
C	11.9	−11.4	3.4	6.1	−5.9	1.6
M	11.5	−5.6	2.0	12.8	−12.5	2.0
CH	16.7	−15.9	7.5	12.7	−8.7	2.2
U	18.7	−18.4	5.2	8.7	−11.3	2.6

[a] D, Dreiding 2.21; C, CFF91; M, MMFF94; C, CHARMM; U, UFF 1.02.

σ_i in eq 1 are 0.001 for energies, 0.01 for gradients, and 0.1 for Hessian matrixes.

Each of the 25 test molecules were fitted and validated independently, which only took a few seconds for each job after the ab initio data was prepared. The calculated bond lengths and angles are compiled in Figure 3, parts a and b, for comparison. Both HF/6-31G(d) and derived force field results are plotted against the reference data in the charts. Clearly, excellent correlations are obtained. Statistical analysis of these results is presented in Table 3. Overall, data calculated using the force fields derived from ab initio data resemble the performance of corresponding ab initio calculations with very similar values in maximum and standard deviations, as illustrated in Table 3.

TABLE 2: Maximum Percentage Deviations of Calculated Bond Lengths and Angles Using Different Force Fields[a]

molecule	sym	D	C	M	CH	U
cyclobutene	C_{2v}	8.2	2.6	1.4	10.3	5.1
methylenecyclopropane	C_{2v}	9.0	4.3	7.0	7.8	11.3
isobutylene	C_s	9.0	2.5	2.4	2.2	3.5
3-methyl-1-butyne	C_s	12.9	5.9	1.8	2.0	3.4
fluorobenzene	C_{2v}	5.6	4.0	1.8	1.1	2.7
aniline	C_{2v}	5.7	1.9	5.3	4.6	4.5
benzonitile	C_{2v}	10.1	3.1	1.5	1.0	1.6
carbonic_difluoride	C_{2v}	11.5	11.5	11.5	1.8	11.5
methylchloroformate	C_s	10.1	5.4	3.0	7.4	5.0
formic anhydride	C_s	10.4	2.9	3.2	4.5	9.9
cinylene carbonate	C_{2v}	4.9	3.0	5.6	16.7	6.6
dimethyl phosphate anion	C_1	14.7	10.7	3.3	15.9	17.2
carboxylate anion	C_{2v}	11.9	3.7	3.5	6.9	9.3
trimethylammonium cation	C_{3v}	0.9	0.9	0.6	2.0	2.6
1,2,5-thiadiazole	C_{2v}	9.7	10.4	3.0	4.5	9.3
1,3,4-thiadiazole	C_{2v}	5.8	4.3	1.1	9.7	3.8
s-tetrazine	D_{2h}	5.6	2.6	1.7	8.2	5.0
pyrazole	C_s	5.3	1.5	1.7	8.2	5.0
pyrrole	C_{2v}	5.2	1.0	1.6	3.1	4.8
2,6-difluoropyridine	C_2v	6.6	5.0	2.4	3.3	5.7
pyrrolidine	C_s	2.9	5.2	1.7	3.8	4.4
morpholine	C_s	3.2	2.6	3.2	3.7	3.4
succinic anhydride	C_{2v}	4.2	2.7	12.8	9.0	6.8
benzenesulfonic acid	C_1	11.9	11.9	3.3	5.1	18.7
benzenesulfonamide	C_s	12.0	9.3	4.6	5.5	17.3

[a] Same as Table 1.

3.3. Sulfonamides. A simple parametrization using an optimized structure has a limited application restricted mostly to rigid molecules (e.g., aromatic rings). Generally speaking, conformational spaces need to be sampled by including distorted structures so that derived force fields can be used to predict a broad range of molecular properties not only for optimized structures but also for any conformational isomers.

Figure 4 illustrates the target compounds to be parametrized and the model compounds (fragments) selected to represent the target molecules. Two to three model compounds were used to represent each of target molecules:

$$(1) = (5) + (6)$$

$$(2) = (5) + (7)$$

$$(3) = (5) + (10) + (11)$$

$$(4) = (5) + (8) + (9)$$

A model compound should be selected so that it contains at least four atomic units. This is because the largest term in a force field, the torsion term, requires four atoms in its definition. If a functional group such as an aromatic ring is located at the end of the fragment, it should be treated as one unit. Fragments may be overlapped. By fitting all fragments simultaneously, optimized parameters should compromise for any variations found in overlapping regions of different fragments.

The size of model molecules is not as important a factor as the number of degrees of freedom in internal rotations. If a model compound contains too many degrees of freedom in internal rotations, the fit can be very difficult not only because

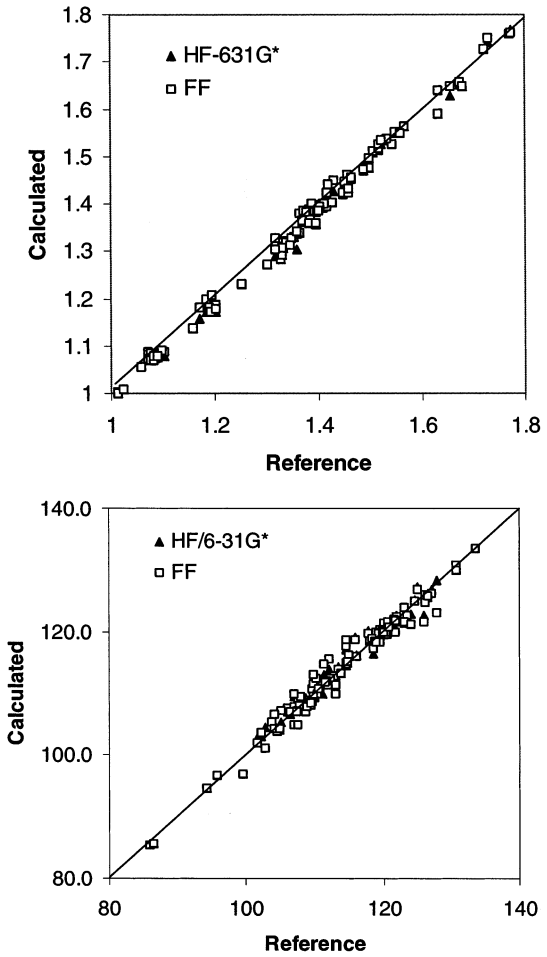

Figure 3. Comparison of calculated and reference bond lengths (a) and angles (b). The calculated data are obtained using the HF/6-31G-(d) method and force field derived from the ab initio data. The bond lengths are in angstrom, and the angles are in degrees.

TABLE 3: Comparison of Percentage Errors in Bond Lengths and Angles Calculated for the Testing Molecules Using HF/6-31G(d) Method and the Force Fields Derived from the ab Initio Data

	bond lengths			bond angles		
	max(+)	min(−)	STD	max(+)	min(−)	STD
HF/6-31G(d)	1.3	−4.0	0.9	2.8	−2.6	1.0
force field	1.5	−3.3	0.9	3.6	−3.8	1.3

too many parameters may be involved but also because the complexity of conformational space makes it difficult to sample correctly. The largest model molecules used in this work probably represents the limit; each of the models 7, 8, and 10 contain three flexible rotation bonds. It should be noted that most parametrization work published so far was based on smaller model compounds where missing parameters are transferred when result parameters are applied to larger molecules.

For molecules that have more than two flexible rotation bonds (usually bonds with single-bond order), a systematic sampling of the conformational space is not preferable because the number of samples can be very large (N^d, where N is the number of sampling points on each of the torsional coordinates and d is the number of the coordinates). Additionally, many of the conformers may have extremely high energies, which makes the fit impractical.

Several procedures have been used to sample the potential energy surfaces in force field developments. One of the methods

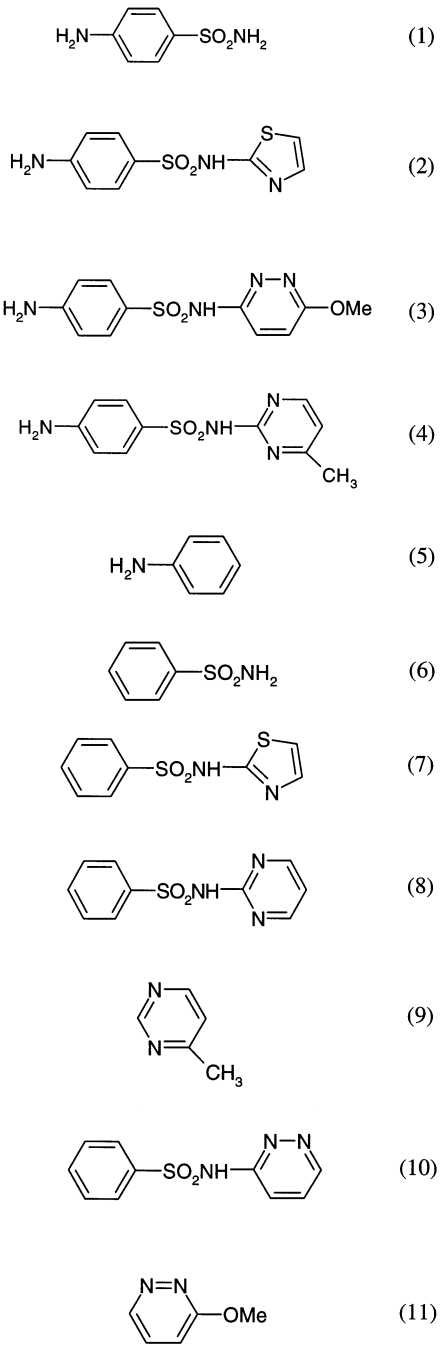

Figure 4. Target and model molecules for parametrization of sulfonamides.

is to randomly distort a number of normal mode coordinates.[7−9] This approach samples multiple internal coordinates simultaneously, which automatically includes couplings among terms in different coordinates. However, it is ideal for small amplitude distortions (e.g., bond stretches and angle distortions) not efficient for large amplitude internal rotations. Another method is to distort one of the internal coordinate while keeping other coordinates frozen at their equilibrium values (we refer this as "adiabatic sampling" in this paper). The "adiabatic sampling" method is associated with the assumption that force field interaction terms are additive, the total energy is a summation of multiple energy terms; each of them is a function of a single internal coordinate. By fixing all other coordinates, one can probe the "intrinsic" energy profile of a particular energy term by distorting the variable coordinate only. It is easy to understand the difference between these two methods on a 2-D

energy contour map such as the well-known $\phi-\psi$ map in peptides. The first method samples data randomly on the map; the second method samples along two perpendicular lines, one is parallel to ϕ (while ψ is fixed) and another is parallel to ψ (while ϕ is fixed). In addition, one could relax other coordinates while fixing one at given values (restrained optimization). This corresponds to a sampling path along the valley on the 2-D contour example. It should be noted that none of these methods samples the surface completely. They are all approximations to replace otherwise too expensive and possibly unnecessary (because the energies may be too high to be useful in force field representation) full scan of the energy surfaces.

In this work, we used the "adiabatic sampling" method which worked reasonably well in previous work.[6−9] The molecules were distorted by rotating the dihedral angle of interest through fixed intervals. For each distorted structure, single-point energy and gradient calculations were performed at the B3LYP/6-311G-(d,p) level of theory.

No gas-phase experimental data were found for sulfonamide functional groups. Therefore, we relied on ab initio calculations to derive and validate force field parameters. For the sake of efficiency, we applied density functional method B3LYP with the 6-311G(d,p) basis set to prepare data. To verify that this method is accurate enough for parametrization purposes, we tested it on three related molecules, aniline, thiazole, and pyridazine, for which there exists high-quality gas-phase experimental data[31] to compare our results with. The calculated bond lengths, angles, and dihedral angles (for aniline only) agree well with the experimental data; the percentage deviations in all bond lengths and angles are less than 2%. (Details are listed in the Supporting Information.) For conformational energies, this method appears to be accurate enough for the purpose. For example, the calculated energy barrier of rotation about C−C bond in ethane is ca. 2.7 kcal/mol, compared favorably with the experimental value of 2.88 kcal/mol.[34]

Because ESP atomic charges obtained are correlated with conformations, we derived charge parameters by essentially fitting to all ESP charges calculated for the model molecules, optimized and distorted structures, in a least-squares sense. However, another problem emerged with these calculations. As reported in the literature, ESP charges obtained for large or bulky molecules tend to be overestimated. For example, the partial charges obtained for nonpolar hydrogen were in the range of 0.15−0.18 electrons, which is significantly larger than the accepted charge value of approximately 0.12.[6] Following the restricted ESP approach, we applied a scaling factor of 0.8 to ESP values to derive charge parameters.

The ab initio energies, gradients and Hessian matrix elements calculated for all model molecules and their distorted structures were used as input data to derive valence parameters by least-squares fit. CFF type functional forms as given in eq 4 and 5 were used in this study. Other options were identical to those delineated in the previous section. There are total of 53 structures used in the fit, which corresponds to 3174 gradient values and 12027 Hessian matrix elements. Note that Hessian matrixes were calculated only for optimized structures. Fit results are illustrated in Figure 5a−h. Energies (5a), gradients (5b), and Hessian elements (5c) were well reproduced, with root of mean squares (RMS) of 0.32 kcal/mol, 3.2 kcal/mol Å, and 13.8 kcal/mol Å^2 respectively.

The conformational energies calculated cover a broad range from ca. 0 to 14 kcal/mol. Excellent fit was obtained (Figure 5a) over the entire range. The agreement was obtained by equally fitting all energies with the same weight, which indicates the functional forms are capable to accurately represent the energy surfaces in the calculated range.

TABLE 4: Summary of Validation Results[a]

	no. of data	max(+)	max(−)	RMS
bond length (A)	134	0.009	−0.025	0.005
bond angle (°)	209	6.6	−4.2	1.2
dihedral angles (°)	273	7.8	−27.9	4.4
nonbond (A)	665	0.59	−0.19	0.14
frequencies (1/cm)	393	77	−106	26

[a] The maximum and root of mean square deviations of structural and vibrational properties between the ab initio and force field results.

TABLE 5: Comparisons of Crystal Cell Parameters and Densities

alpha	*A*	*B*	*C*	*D*(g/cm^3)
expt	5.650	18.509	14.794	1.479
min	5.730	17.137	14.650	1.590
MD	5.935	17.751	15.175	1.431
dev(%)	5.0	−4.1	2.6	−3.2

beta	*A*	*B*	*C*	beta(deg)	*D*(g/cm^3)
expt	8.975	9.005	10.039	111.430	1.514
min	8.806	8.854	9.990	115.641	1.629
MD	9.047	9.096	10.263	115.641	1.502
dev(%)	0.8	1.0	2.2	3.8	−0.8

gamma	*A*	*B*	*C*	beta(deg)	*D*(g/cm^3)
expt	7.950	12.945	7.790	106.500	1.486
min	7.767	12.629	7.639	104.479	1.577
MD	7.929	12.892	7.798	104.479	1.482
dev(%)	−0.3	−0.4	0.1	−1.9	−0.3

The derived parameters were validated by applying them in the optimization of model molecules. The calculated structural parameters and vibrational frequencies are compared with the ab initio B3LYP/6-311G(d,p) results in Figure 5d−h. Statistical analyses of these comparisons are listed in Table 4. The RMS deviations are 0.005 Å for bond lengths, 1.2° for bond angles, 4.4° for dihedral angles, 0.14 Å for nonbond distances, and 26 cm^{-1} for normal-mode frequencies. The results are generally satisfactory, provided that some model molecules are fairly large and contain three flexible internal rotation bonds. In particular, the −SO2−NH− group is difficult to parametrize because the nitrogen center is puckered based on ab initio calculations while remaining extremely flexible in terms of inversion and rotation. It is well accepted that the amino group is planar in crystals because of a packing effect. The largest deviations (6.6 degrees in bond angles, and 27.9 degrees in dihedral angles) are related to the N−H bond.

By fitting to conformational energies, the resulting force field is capable of being used in a study of conformational properties. Figure 6 illustrates optimized rotational energy profiles calculated for sulfanilamide using the result force field and ab initio (B3LYP/6-311G(d,p)) method. Both force field and ab initio calculations were restrained optimization. It is of interest to note that the force field reproduces the optimized torsion profiles well, although the parameters were derived based on "adiabatic sampling" as explained above.

Because of flexible rotational bonds and polarized atoms, conformational variations of sulfonamides appear to be quite complicated. It is of interest to note that the rotation about the C(ar)−S bond is very flexible, with the barrier height of rotation being only ca. 2 kcal/mol. On the other hand, a rotation about the S−N bond is apparently influenced by the intramolecular hydrogen bonds between the sulfonate oxygen and the amino hydrogen (S═O···H−N) and shows asymmetric behavior with

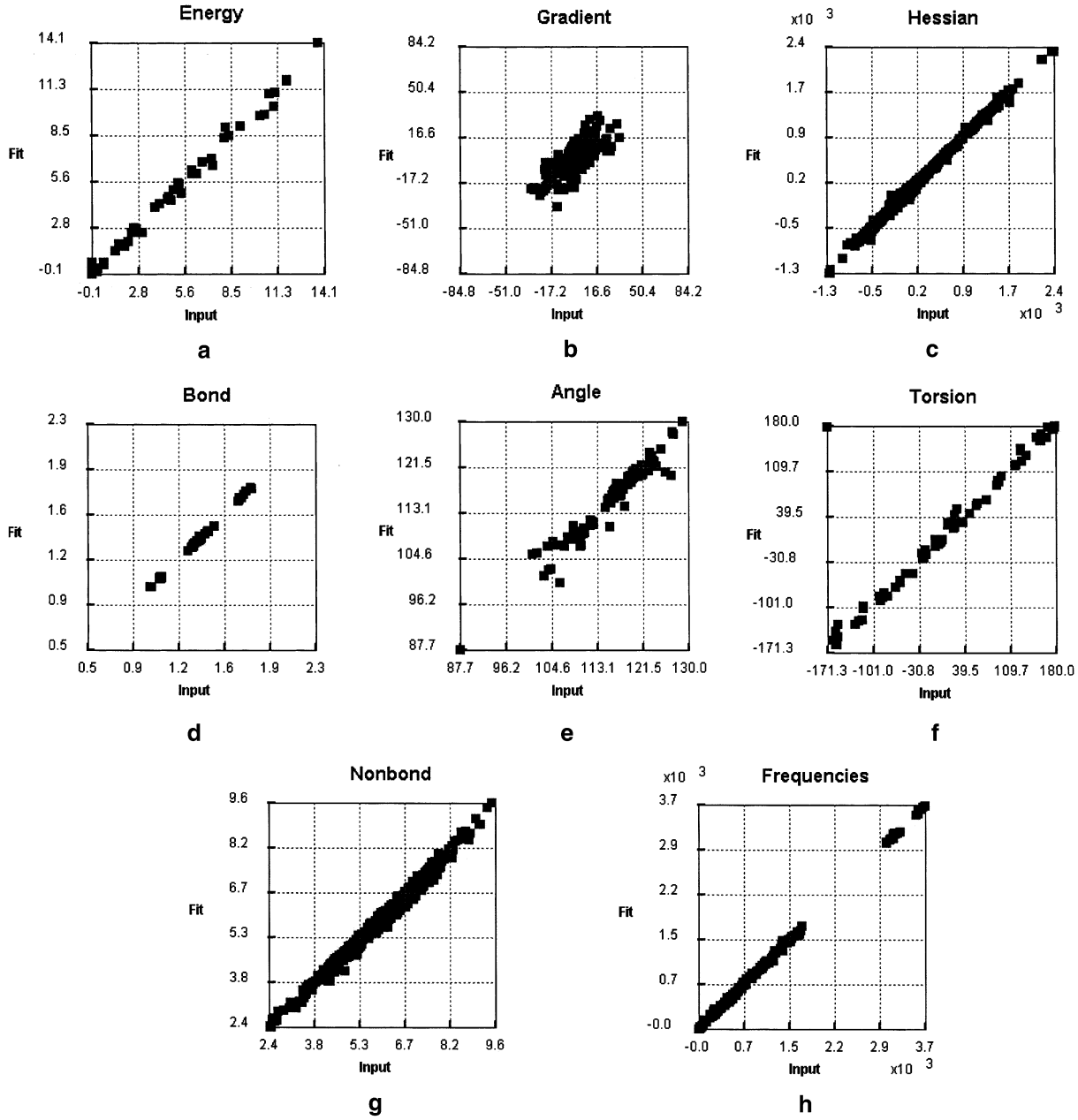

Figure 5. Comparisons of fit and validation results of force field parametrization for sulfonamides. (a) total energies (in kcal/mol), (b) gradients (in kcal/mol Å), (c) Hessian matrix elements (in kcal/mol Å^2), (d) bond lengths (in Å), (e) bond angles (in degrees), (f) torsion angles (in degrees), (g) nonbond distances (in Å), and (h) normal model frequencies (in cm^{-1}).

respect to the dihedral angle of C−S−N−H in Figure 6b. The barrier height of rotation is ca. 3−4 kcal/mol.

The nonbond VDW parameters were fixed to a set of default values taken from literature when valence parameters were derived. Following our previous work,[6] these parameters were then validated and optimized using condensed phase simulations. Three stable crystalline forms, alpha, beta, and gamma, have been reported for sulfanilamide in the literature.[35−37] We carried out molecular dynamics simulations using these crystal structures.

We first performed energy optimization using for each crystal unit cell to verify that optimized molecular structures are in good agreement with the crystalline data. The standard percentage errors between the force field optimized bond lengths and the reported data for three crystalline forms are between 1 and 2.1%, whereas the maximum percentage error is less than 4.9%. Smaller percentage errors are obtained for the bond angles. Three different values, 59.0, 109.4, and 88.0°, are reported for the dihedral angle C−C−S−N in the alpha, beta, and gamma forms, respectively. Our calculated values are 61.3, 115.8, and 92.4°.

Using both energy minimization and MD simulation techniques, we obtained an overall agreement of crystal cell parameters and densities by adjusting only the VDW parameters of the amino hydrogen from 1.02 to 1.82. Other parameters were tested and found to be insensitive to the properties of interest. The final results of the calculations are listed in Table 5. As expected, there are systematic differences between minimization and dynamics results. With MD simulations, we found that crystal cell parameters and densities are in good agreement with experimental data, whereas energy minimization overestimated densities by ignoring thermal expansion.

The relative total energies of the simulated three forms of crystal sulfanilamide are 0.0, −0.7, and +19.0 kcal/mol per

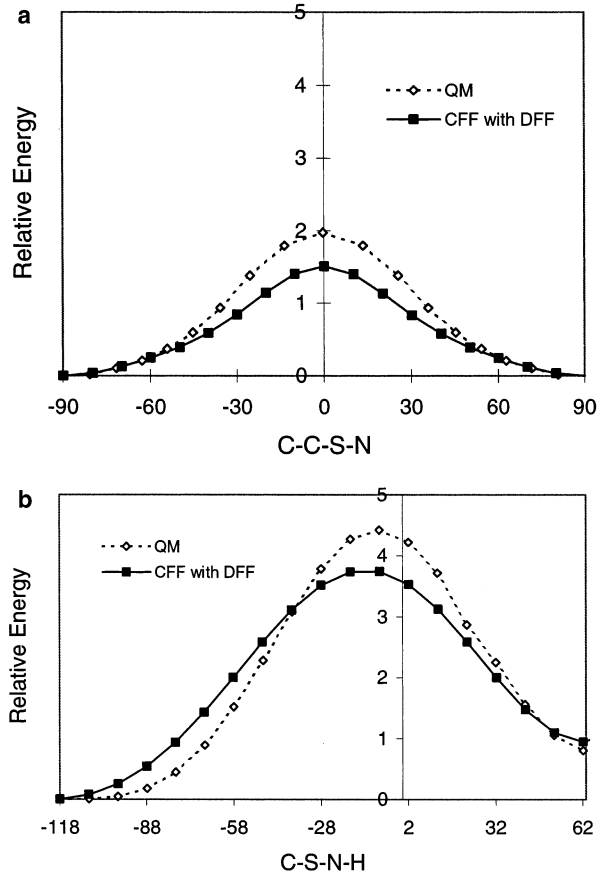

Figure 6. Comparison of torsion profiles optimized using the derived force field for sulfonamides. (a) Profile of dihedral angle of C−C−S−N, and (b) profile of dihedral angle of C−S−N−H. The energies are in kcal/mol, and angles are in degrees.

molecule, with estimated uncertainty of ca. 0.2 kcal/mol. Assuming the intramolecular energies are the same in these forms, the calculated relative stability of these forms is in line with the experimental observations.[38,39]

After modifying the VDW parameters, we tested valence parameters again by repeating the fit procedure. The impact to valence parametrization is negligible. Final parameters are listed in the Supporting Information.

Conclusion

Although it is well-known that a force field only works for what it is designed for,[1−3] how parameters are transferred and what the consequences of parameter transferal are discussed in this paper. We analyzed the problems associated with transferring parameters: parameters are used for molecules that are not parametrized. Our conclusion is that once parameters are transferred externally the errors in results calculated using a parametrized force field are similar to those obtained using a generic one, which could be greater than 15% in structural parameters. To obtain high accuracy in calculations using force fields, especially those specifically parametrized, validation must be applied if the intend use is outside of the parametrization scope.

Because the chemistry diversity is so large, it is unlikely a force field can be made to meet every need. A natural solution to these problems is to parametrize force field specifically whenever it is necessary. We investigated this approach by using automatic parametrization software. The results obtained are promising. Highly accurate force field parameters can be derived rapidly from ab initio data, and the predicted molecular properties are essentially reproductions of the corresponding ab initio results.

The parametrization procedure was applied to sulfonamide molecules in order to prepare a set of high-quality parameters for future studies and to test how this approach can be applied to realistic drug molecular systems. A group of target molecule-derived fragments, representing all required interaction terms, were used in this study. By fitting all model molecules simultaneously, the parameters obtained accurately describe potential energy surfaces for these molecules and yield good agreements with ab initio results in structures, conformational energies and vibrational frequencies. The resulting force field was applied in simulations of crystal structures to validate and refine the nonbonded VDW parameters. Using this example, we demonstrated how a rapid parametrization of force field from ab initio data could provide a solution to the problems associated with transferring parameters. Furthermore, this approach allows force field methods to be directly based on quantum mechanics calculations. When parameters can be derived from ab initio data rapidly and accurately, as illustrated in this paper, force field calculations no longer need to be empirical and thus limited by existing data. Instead, they become scale-up extensions of quantum mechanics calculations that can be applied to potentially any molecular systems, existing or not.

Supporting Information Available: Calculated bond lengths, angles, and dihedral angles compared with the experimental data. This material is available free of charge via the Internet at http//pubs.acs.org.

References and Notes

(1) Burkert, U.; Allinger, N. L. *Molecular Mechanics, ACS Monograph 77*; Caserio, M. C., Ed.; American Chemical Society: Washington, DC, 1989.

(2) Bowen, J. P.; Allinger, N. L. In *Reviews in Computational Chemistry*; Lipkowitzand, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1992; Vol. 2, p 81.

(3) Dinur, U.; Hagler, A. T. In *Reviews in Computational Chemistry*; Lipkowitzand, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1992; Vol 2, p 99.

(4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(5) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M., *J. Am. Chem. Soc.*, **1995**, *117*, 11946.

(6) Sun, H. *J. Phys. Chem. B.* **1998**, *102*, 7338.

(7) Maple, J. R.; Dinur, U.; Hagler, A. T. *Proc. Natl. Acad. Sci*. **1988**, *85*, 5350.

(8) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Comput. Chem.* **1994**, *15*, 162.

(9) Sun, H.; Mumby, S. J.; Maple, J. R.; Hagler, A. T. *J. Am. Chem. Soc.* **1994**, *116*, 2978.

(10) Halgren, T. A. *J. Comput. Chem*. **1996**, *17*, 490.

(11) Dasgupta, S.; Goddard, W. A. *J. Chem. Phys.* **1989**, *90*, 7207. Dasgupta, S.; Yamasaki, T.; Goddard, W. A. *J. Chem. Phys.* **1996**, *104*, 2898.

(12) *Direct Force Field*; Aeon Technology, Inc.: San Diego, CA, 2002.

(13) Gringauz, A. *Introduction to Medicinal Chemistry, How Drugs Act and Why*; Wiley-VCH: New York, 1996.

(14) Kiyama, R.; Tamura, Y.; Watanabe, F.; Tsuzuki, H.; Ohtani, M.; Yodo, M. *J. Med. Chem*. **1999**, *42*, 1723.

(15) Nicholas, J. B.; Vance, R.; Martin, E.; Burke, B. J.; Hopfinger, A. J. *J. Phys. Chem.* **1991**, *95*, 9803.

(16) Liang, G.; Bays, J. P.; Bowen, J. P. *J. Mol. Struct. (THEOCHEM)* **1997**, *401*, 165.

(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.;

Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(18) Roothan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.

(19) (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257. (c) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209. (d) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163. (e) Hariharan, P. C.; Pople, J. A. *Theo. Chim. Acta* **1973**, *28*, 213.

(20) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum Chem. Symp.* **1977**, *11*, 149.

(21) (a) Mclean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639. (b) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650. (c) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033. (d) Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377. (e) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.

(22) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(23) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(24) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(25) Cerius^2 Modeling Environment, release 4.5; Accelrys Inc.: San Diego, 2001.

(26) Molecular Operating Environment, version 2001.01; Chemical Computing Group: Montreal, 2001.

(27) MS Discover, release 2001.1; Accelrys Inc.: San Diego, 2001.

(28) (a) Tosi, M. P. *Solid State Phys.* **1964**, *16*, 107. (b) Ewald, P. *P. Ann. Phys.* **1921**, *64*, 253.

(29) Jorgensen, W. L.; Maxwell, D. S.; Tiradorives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(30) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.

(31) Harmony, M. D.; Laurie, V. W.; Kuczkowski, R. L.; Schwendeman, R. H.; Ramsay, D. A.; Lovas, F. J.; Lafferty, W. J.; Maki, A. G. *J. Phys. Chem. Ref. Data.* **1979**, *3*, 619. (This paper contains compiled gas-phase data of polyatomic molecules determined by spectroscopic methods with original paper referenced cited.)

(32) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.

(33) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

(34) Hirota, E.; Endo, Y.; Saito, S.; Dunca, J. L. *J. Mol. Spectrosc.* **1981**, *89*, 285.

(35) O'Connor, B. H.; Maslen, E. N. *Acta Crystallogr.* **1965**, *18*, 363.

(36) O'Connell, A. M.; Maslen E. N. *Acta Crystallogr.* **1967**, *22*, 134.

(37) Alléaume, M.; Decap, J. *Acta Crystallogr.* **1965**, *19*, 934.

(38) Toscani, S.; Thoren, S.; Agafonov, V.; Ceolin, R.; Dugue, J. *Pharm. Res.* **1995**, *10*, 1453.

(39) Toscani, S.; Agafonov, V.; Dugue, J.; Ceolin, R. *Pharm. Res.* **1996**, *1*, 151.