Evaluation of Computational Chemistry Methods: Crystallographic and Cheminformatics Analysis of Aminothiazole Methoximes

Tulay Ercanli and Donald B. Boyd*

Department of Chemistry, Indiana University−Purdue University at Indianapolis, 402 North Blackford Street, Indianapolis, Indiana 46202-3274

Received October 30, 2004

Cheminformatics is used to validate the capabilities of widely used quantum chemistry and molecular mechanics methods. Among the quantum methods examined are the semiempirical MNDO, AM1, and PM3 methods, Hartree−Fock (ab initio) at a range of basis set levels, density functional theory (DFT) at a range of basis sets, and a post-Hartree−Fock method, local Møller−Plesset second-order perturbation theory (LMP2). Among the force fields compared are AMBER*, MMFF94, MMFF94s, OPLS/A, OPLS-AA, Sybyl, and Tripos. Programs used are Spartan, MacroModel, SYBYL, and Jaguar. The test molecule is (2-amino-5-thiazolyl)-alpha-(methoxyimino)-*N*-methylacetamide, a model of the aminothiazole methoxime (ATMO) side chain of third-generation cephalosporin antibacterial agents. The Ward hierarchical clustering technique yields an insightful comparison of experimental (X-ray) and calculated (energy optimized) bond lengths and bond angles. The computational chemistry methods are also compared in terms of the potential energy curves they predict for internal rotation. Clustering analysis and regression analysis are compared. The MMFF94 force field such as implemented in MacroModel is the best overall computational chemistry method at reproducing crystallographic data and conformational properties of the ATMO moiety. This work demonstrates that going to a higher level of quantum theory does not necessarily give better results and that quantum mechanical results are not necessarily better than molecular mechanics results.

INTRODUCTION

Selection of the most appropriate computational chemistry method for modeling a given system is obviously crucial for obtaining reliable results. It is hard to know a priori what level of quantum theory or which molecular mechanics force field is appropriate without a careful comparison of the methods in regard to their ability to reproduce relevant experimental data. We want to know which method is best at modeling a pharmaceutically related, medium-sized, organic molecule. For our purposes, the best method should give optimized geometrical variables closest to experimental geometrical variables and potential energy curves suitable for conformational analysis.

Aminothiazole methoxime (2-amino-4-thiazolyl-methoxyimino, ATMO) (Figure 1) is a substructure of the 7-acylamido side chain of well-known, marketed, third-generation cephalosporin antibacterial agents, such as cefotaxime, ceftizoxime, cefmenoxime, ceftriaxone, and ceftazidime.[1] In addition, ATMO has been incorporated in literally thousands of cephalosporins that have synthesized at pharmaceutical companies during investigations of structure−activity relationships.[2,3] The ATMO substructure is also of interest because it is a component of other biologically active materials.[4]

In the present study, we investigate the reliability of an extensive variety of popular computational chemistry methods in terms of their ability to handle accurately the ATMO moiety. The test molecule is challenging because it has some unusual combinations of functional groups not in the usual repertoire of computational chemistry test sets.

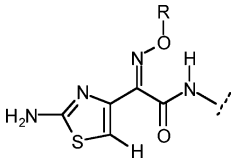

Figure 1. ATMO substructure, R = methyl.

Benchmarking, calibrating, and validating computational chemistry methods have always been important to advancing the field.[5−7] A few selected examples of such studies include the following. Liljefors et al.[8,9] compared the performance of 14 force fields, as well as a small selection of semiempirical, Hartree−Fock (HF), and density functional theory (DFT) methods in terms of their ability to reproduce experimental conformational energy data for a variety of small organic compounds including acyclic and cyclic hydrocarbons, as well as oxygen-containing and nitrogen-containing compounds. MM2* and MMFF results were found to be comparable to DFT results and significantly more successful than HF results. The semiempirical molecular orbital method PM3 had the least accuracy. A study by Alagona et al. indicates that MMFF94 is the right choice when ab initio and DFT calculations are not affordable, and this force field is quite accurate for pterocarpan structure analysis.[10] Indeed, the results of the study of Blatchly and co-workers showed that MMFF is very good for helical systems and helix formation energy calculations; the authors concluded that MMFF is better than AM1 and DFT.[11]

* Corresponding author phone: (317)274-6891; fax: (317)274-4701; e-mail: boyd@chem.iupui.edu.

10.1021/ci049671x CCC: $30.25
Published on Web 04/02/2005

The usual way to compare experimental and theoretical calculated properties is to compute root-mean-square deviations, which give a precise quantitation of agreement (or lack thereof). Cluster analysis of cheminformatics fame is another approach for comparing data.[12] We explore its use here and, as a result, advocate further utilization of this appealing approach. A characteristic of cluster analysis is that many properties can be compared simultaneously and globally. Each property is represented as one dimension in multidimensional space. For example, each bond length, each bond angle, or each other molecular property would be one dimension of the study space. The results from cluster analysis are quickly presented visually in terms of a tree diagram called a dendrogram. In comparing computational chemistry methods, for instance, those that give similar predictions of molecular properties would be presented on tree branches that are close together. In contrast, methods that give widely different predictions appear on widely separated branches of the tree.

Here we compare data from 27 computational chemistry methods. An earlier study has been published in which cluster analysis was used to compare the performance of a few quantum chemical methods applied to rhodanine, a pharmaceutically related ring system.[13]

METHODOLOGY

The Cambridge Structural Database (CSD) is a collection of molecular structure data derived from X-ray and neutron diffraction experiments on more than 300 000 organic and organometallic compounds.[14] Fourteen compounds in the CSD contain ATMO as a substructure.[15−27] Three-dimensional atomic coordinates for one hit, CITZUA, are not available. Another hit (TAWYAR) has three iron ions in its crystal structure; this hit was excluded from analysis because the complexation could induce changes in the ATMO structure.

JMP[28], a versatile, interactive statistics program, was used to perform distribution analysis on the experimental bond lengths and bond angles. The resulting histograms revealed the frequency of occurrence of specific bond length or bond angle values in the ATMO-containing compounds. JMP was also used to compute mean values, standard deviations, and standard errors. Crystalline state dihedral angles at three rotatable bonds in the ATMO compounds were examined. These rotatable bonds are the ones relevant to conformational variation among ATMO-containing structures. Torsional angles observed in the crystalline compounds are visualized using wheel diagrams (0° to 360°) drawn by AutoCad.[29]

For our computational chemistry calculations, Spartan[30] software was run on both PC (running Windows) and Silicon Graphics workstations (running UNIX, www.sgi.com). SYBYL,[31] MacroModel,[32] and Jaguar[33] software were run on Silicon Graphics workstations. To expedite the numerous calculations performed, we used a farm of workstations running simultaneously.

The starting structure for the molecular modeling was built in Spartan (Figure 2). A methyl group was added to the amide nitrogen of the ATMO substructure to make the structure chemically complete. Thus, the model treated was (2-amino-5-thiazolyl)-alpha-(methoxyimino)-*N*-methylacetamide. (This calculated structure has CAS Registry Number 752165−96−1; synonym: (*Z*)-2-amino-alpha-(methoxyimino)-*N*-methyl-4-thiazoleacetamide.) Optimized bond lengths and bond angles were determined by energy minimization in each method (with defaults in each program). Next, single-point energy calculations are done by changing the dihedral angles at the three rotatable bonds in our model structure. In each case, these rotatable bonds were varied in increments while bond lengths, bond angles, and the rest of the dihedral angles are kept fixed at their previously optimized values.

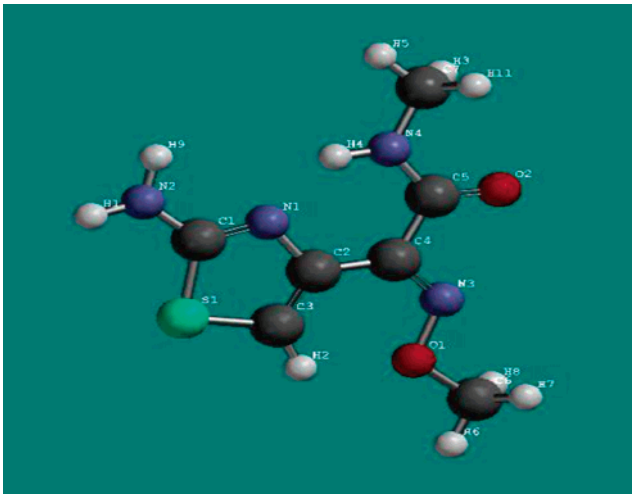

Figure 2. Structure used in computational chemistry calculations. The atom numbering system used in this paper is shown.

Semiempirical[34,35] MNDO, AM1, and PM3 molecular orbital calculations were carried out in Spartan. Spartan was also used for gas-phase HF and DFT calculations with a variety basis sets ranging from small to large: HF/ STO-3G, 3-21G(*), 6-31G*, 6-31G**, 6-31+G*, and 6-311G*; DFT/ 6-31G*, 6-31G**, 6-31+G*, 6-311G* and 6-311+G**.[6,7,36−38] The DFT calculations were run with the B3LYP[39] functional.

Molecular mechanics was run with Spartan's MMFF94[40−45] and Sybyl force fields. The latter is an implementation of the Tripos[46] force field. (Consistent with software manuals, we use Sybyl to indicate the force field and SYBYL to indicate the well-known molecular modeling program.) In SYBYL, molecular mechanics calculations were done with the Tripos, MMFF94, and MMFF94s[47] force fields. The MMFF94, MMFF94s, AMBER*,[48] OPLS/A,[49] and OPLS-AA[50] force fields were used as implemented in MacroModel (abbreviated as MMod in the tables and figures of this paper). The Tripos force field was developed for organic and bioorganic molecules in general. MMFF94 is designed for modeling organic molecules, especially those of pharmaceutical interest. The difference between MMFF94 and MMFF94s is that the latter is better at yielding planar delocalized sp^2 nitrogens instead of pyramidal ones.[47] AMBER was developed primarily for large proteins and nucleic acids,[51] and AMBER* is the implementation in MacroModel. OPLS/A is a force field intended for use mainly on biopolymers and carbohydrates especially in solution. OPLS-AA works well for condensed-phase simulation of peptides. It can be suitable to use on some drug-like molecules and is expected to work better than OPLS/A because of its better parameterization.[32,50] All these force fields are all-atom force fields meaning hydrogens are treated explicitly.

Table 1. ATMO-Containing Compounds in the Cambridge Structural Database along with Their Crystallographic *R* Value

hit	chemical name	*R*	ref
BAGSAD	2-(2-amino-4-thiazolyl)-2(*E*)-(methoxy)iminoacetic acid	0.08	15
BAGSEH1	2-(2-amino-4-thiazolyl)-2(*Z*)-(methoxy)iminoacetic acid trihydrate	0.038	16
BAGSEH2	2-(2-amino-4-thiazolyl)-2(*Z*)-(methoxy)iminoacetic acid trihydrate	0.038	16
CFTXIM	2-(2-amino-4-thiazolyl)-2(*Z*)-(methoxy)imino ethyl acetate	0.041	18
EAMTHZ	ethyl 2-amino-alpha-(*E*-methoxyimino)-4-thiazoleacetate hydrobromide	0.048	19
EATZAC	ethyl 2-amino-alpha-(*E*-methoxyimino)-4-thiazoleacetate	0.074	20
FAJMAE	(6R,7R)-7-((*Z*)-2-(2-amino-4-thiazolyl)-2-methoxyiminoacetamido)-8-oxo-5-thia-1-azabicyclo(4.2.0)oct-2-ene-2-carboxylic acid	0.074	21
FAJMEI	(6R,7R)-7-((*Z*)-2-(2-amino-4-thiazolyl)-2-methoxyiminoacetamido)-8-oxo-5-thia-1-azabicyclo(4.2.0)oct-2-ene-2-carboxylic acid hydrochloride monohydrate	0.076	21
FUPKOQ	1-((2-amino-4-thiazolyl)(methoxyimino)acetyl)-1H-benzotriazole-3-oxide	0.04	22
GAFTIQ	2-(2-amino-4-thiazolyl-5-chloro)-2-*Z*-methoxyimino-*N*-methylacetamide	0.042	23
GOHRAW	2-(2-amino-4-thiazolyl)-2*Z*)-methoxyiminomethylacetamide hydrochloride monohydrate	0.078	24
VENWIU	7-[(*Z*)-2-(2-aminothiazol-4-yl)-2-(imidazol-4-ylmethoxyimino)acetamido]-3-[(1-pyridinio)methyl]-3-cephem-4-carboxylate diperchlorate dehydrate	0.0938	25

Local Møller−Plesset second-order perturbation theory[52] (LMP2) calculations were done in Jaguar at the 6-31G* level. The effect of solvation on the model structure was modeled by the self-consistent reaction field method[53] (SCRF) as implemented in Jaguar. SCRF was run at both the HF 3-21G-(*) and DFT 6-31G* levels of theory.

To avoid overweighting any of the dimensions, all the data in a cluster analysis should be of the same magnitude. Another caveat in regard to using cluster analysis is that it can be very sensitive to changes in the data set. This sensitivity leads to the fact that errors in the data set can seriously affect the outcome of the analysis. However, visual presentation of the output from cluster analysis often makes it easy to spot such problems.

Our data set of results was assembled as follows. Bond lengths were expressed in angstroms, and they were in the 1.1−1.8 Å range. Bond angles were between 86 and 130°. To make the distance and angle values of closely comparable magnitude, we converted the angles to radians (1 rad = 180/3.14 degrees). The resulting bond angle range is 1.5−2.3 rad. Similarly, we converted the computed molecular energies from kcal/mol or hartrees to units of electron volts (1 kcal/mol = 0.04338 eV; 1 hartree = 27.22 eV) to make them also of closely comparable magnitude. Due to the way we truncate conformational energies as explained later, the resulting energy range is 0.0−1.1 eV. Hence all data are naturally scaled to fall in the range 0.0−2.3.

Data sets containing calculated and experimental data were analyzed by JMP and processed with the Ward hierarchical agglomerative clustering technique.[12] This clustering routine first identifies the most homogeneous subgroups in the data set. Then these subgroups are successively gathered into larger and larger groups until the entire data set is organized hierarchically. The subgroups are agglomerated based on distances between them in multidimensional space. Dendrograms graphically present the clustering results. The length of the tree branches in the dendrogram indicates the proximity between variables in the study space examined. This approach allows us to readily see the relatedness of the computational methods with each other and with the experimental data.

RESULTS AND DISCUSSION

The optimized three-dimensional structure of the ATMO model system is studied using 10 molecular mechanics implementations and 17 levels of theory in quantum chemistry calculations. The degree of similarity between experimental and computational data allows us to identify the best method or methods.

Distribution curves for all bond lengths and bond angles were scrutinized. The point of this check is to look for data abnormally deviating from the rest. One hit (BODKOU) was an outlier in all distribution plots and was excluded from further study. The X-ray determination of BODKOU had only a limited number of usable reflections and a very poor final *R* value of 0.123.[17] The 11 remaining compounds used in this work are listed in Table 1 along with their experimental crystallographic *R* values. One compound (BAGSEH) occurs with two independent molecules per unit cell, so this gives us two sets of atomic coordinates. Hence there are a total of 12 three-dimensional structures for our analysis.

Experimental bond lengths (Table S1; see Supporting Information) and bond angles (Table S2) of the ATMO-containing compounds are obtained from the CSD and used to calculate mean values for each internal geometrical variable. These means are used in the cluster analysis as the experimental values. In Tables S1 and S2 we also show standard deviations and standard errors. Experimental dihedral angles of these compounds are listed in Table S3. Theoretically determined bond lengths (Table S4) and bond angles (Table S5) were obtained by energy minimization using the various methods already mentioned. Tables S6-S8 present the energies from the single-point calculations on the conformers corresponding to the three rotatable bonds.

The dendrogram in Figure 3 presents the relationship between the mean experimental bond lengths of the ATMO substructure and the calculated values from the 27 computational chemistry methods. Clustering indicates that the relatively inexpensive quantum chemistry method HF 3-21G-(*) comes closest to matching the mean experimental bond lengths. The good performance of HF 3-21G(*) for bond lengths seems almost happenstance because larger basis sets do not give as good results. But given the available experimental data and given the complex variety of atom types encountered in the ATMO substructure, this is the result obtained. Including the effect of solvation in the calculations did not significantly improve the optimized bond lengths. Beyond HF 3-21G(*) and HF 3-21G(*)/SCRF, the cluster next closest to the experimental data contains the various MMFF results. In contrast, the Sybyl, Tripos, and

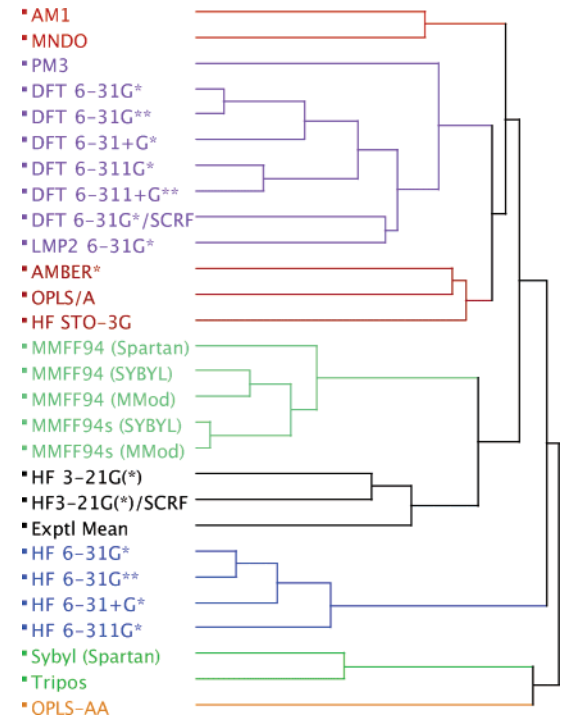

Figure 3. Dendrogram obtained using JMP for analysis of bond lengths.

OPLS-AA force fields yield the poorest optimized bond lengths. Hartree−Fock calculations with basis sets larger than 3-21G(*) are almost as poor as these three force field implementations. Semiempirical and DFT methods are intermediate in their ability to reproduce the experimental bond lengths.

We can further scrutinize these findings by regression analysis (Figure 4). The HF 3-21G(*) correlates best with experimental bond lengths of the test molecule ($r^2 = 0.9903$). Adding the effect of solvation (SCRF) did not alter the outcome significantly ($r^2 = 0.9930$). The various implementations of the MMFF94 force field perform well; for instance, the implementation in MacroModel shows a good fit ($r^2 = 0.9912$). Regression analysis plots show, of course, a line best fitting the data. The inner pair of dotted red curves show the 95% confidence limits of the line. The outer pair of dotted red curves show the 95% confidence limits for individual data points; any point falling outside of these latter limits can be considered an outlier. In the figure captions, we report standard statistical output from the regression analyses. The *t* ratios are for the intercept and coefficient of the regression equation; an absolute value greater than 2.0 is a common rule of thumb for judging significance of the fit because it approximates the 0.05 significance level. As usual, *p* is the probability of the null hypothesis being satisfied, i.e., chance correlation.

We point out that there is not a strict one-to-one correspondence between the closeness of objects in the dendrogram and the correlation coefficients. This is because clustering involves the proximity of points in multidimensional space, whereas regression analysis involves fitting data to a line. In general, however, those methods that correlate well with experiment show better r^2 values. We explore this correspondence further in Figure 5 where we show clustering of the r^2 values obtained from regressing the

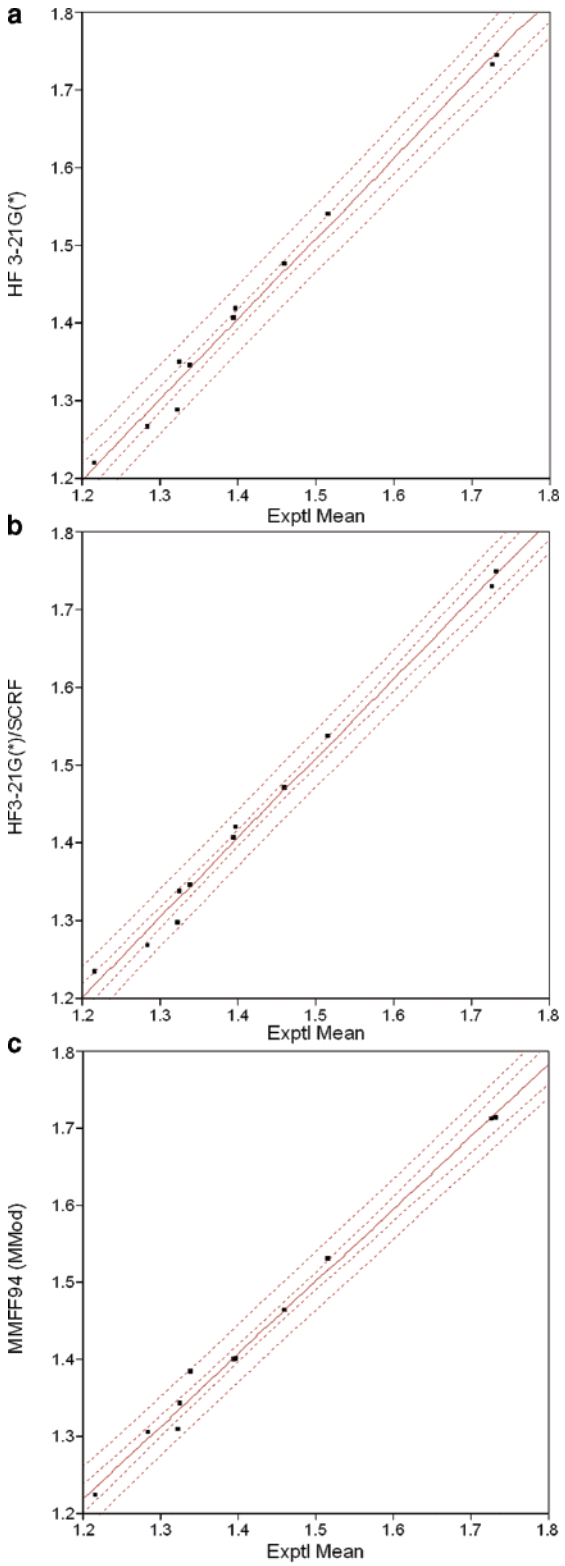

Figure 4. (a) Bivariate fit of HF 3-21G(*) vs experimental mean of bond lengths. Regression equation: HF 3-21G(*) = −0.0418 (±0.0490) + 1.0330 (±0.0340) exptl mean; $n = 11$; $r^2 = 0.9903$; $r_{adj}^2 = 0.9892$; RMS error = 0.0182; *t* ratios = −0.85, 30.34; $p < 0.0001$. (b) Bivariate fit of HF 3-21G(*)/SCRF vs experimental mean of bond lengths. Regression equation: HF 3-21G(*)/SCRF = −0.0250 (±0.0412) + 1.0218 (±0.0287) exptl mean; $n = 11$; $r^2 = 0.9930$; $r_{adj}^2 = 0.9922$; RMS error = 0.0154; *t* ratios = −0.61, 35.63; $p < 0.0001$. (c) Bivariate fit of MMFF94 (in MacroModel) vs experimental mean of bond lengths. Regression equation: MMFF94 (MMod) = 0.0930 (±0.0425) + 0.9390 (±0.0296) exptl mean; $n = 11$; $r^2 = 0.9912$; $r_{adj}^2 = 0.9902$; RMS error = 0.0158; *t* ratios = 2.19, 31.76; $p < 0.0001$.

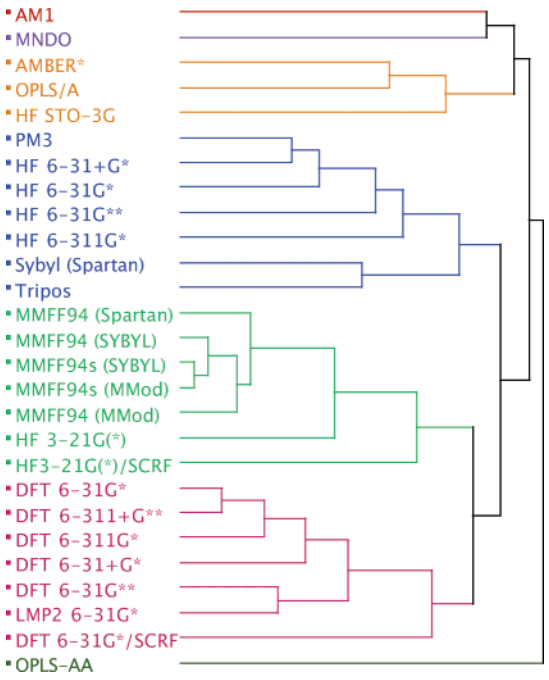

Figure 5. Dendrogram showing the clustering of the r^2 statistics for the fit between experimental mean and optimized bond lengths as predicted by the computational chemistry methods.

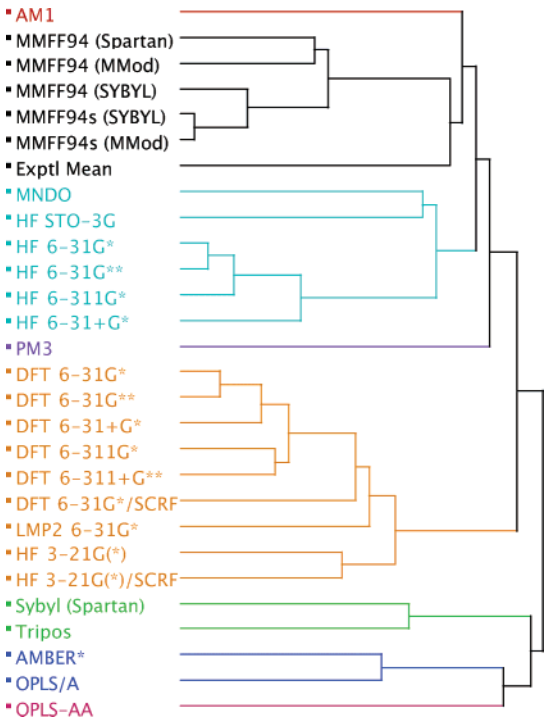

Figure 6. Dendrogram obtained using JMP for analysis of bond angles.

experimental bond lengths and predictions by each of the methods. Note that the hierarchical clustering is only roughly similar to that in Figure 3. However, the DFT results still cluster together, the HF results from the better basis sets still cluster together, and the various MMFF results still form a tight cluster.

The dendrogram in Figure 6 shows the relationship between experimental mean bond angles and calculated values. We see that the MMFF94s force field in SYBYL has the best agreement with the experimental bond angles.

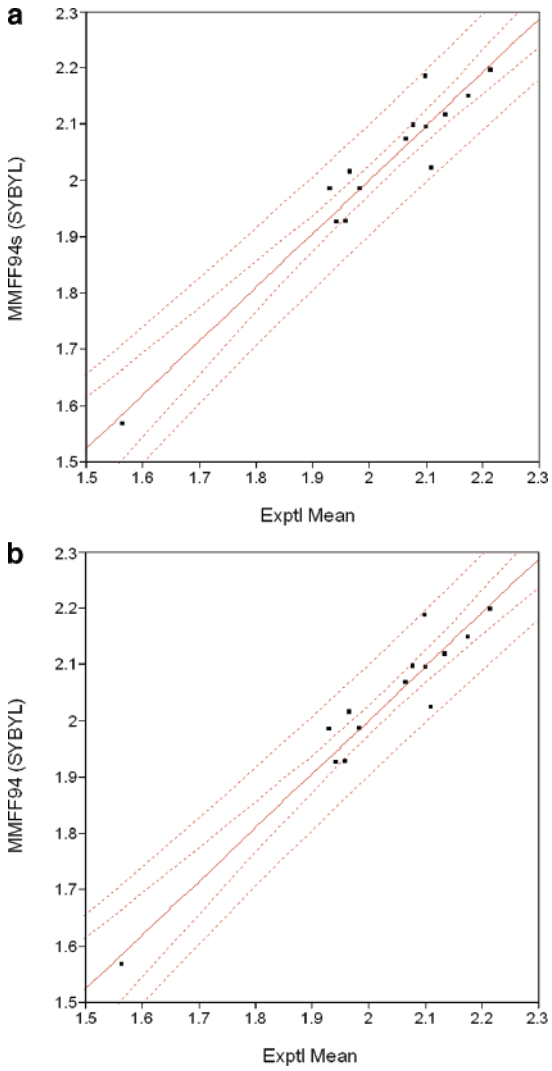

Figure 7. (a) Bivariate fit of MMFF94s (in SYBYL) vs experimental mean of bond angles (in radians). Regression equation: MMFF94s (SYBYL) = 0.0946 (±0.1538) + 0.9534 (±0.0758) exptl mean; $n = 14$; $r^2 = 0.9295$; $r_{adj}^2 = 0.9237$; RMS error = 0.0437; t ratios = 0.61, 12.58; $p < 0.0001$. (b) Bivariate fit of MMFF94 (in SYBYL) vs experimental mean of bond angles. Regression equation: MMFF94 (SYBYL) = 0.0937 (±0.1544) + 0.9537 (±0.0761) exptl mean; $n = 14$; $r^2 = 0.9291$; $r_{adj}^2 = 0.9232$; RMS error = 0.0438; t ratios = 0.61, 12.54; $p < 0.0001$.

Results of MMFF94 in SYBYL also show a close fit to the experimental values. The amount of variance explained and the t ratios in the bivariate fits of these results (Figure 7) indicate the goodness of the fits. For bond angles, all the force fields other than MMFF94 and MMFF94s perform poorly. Density functional theory generally is not as good as Hartree−Fock. The post-Hartree−Fock calculations cost substantially more in computer time but without any compensating improvement in results.

Thus, prediction of bond lengths in the ATMO substructure is quite accurate by HF 3-21G(*), whereas MMFF94 and MMFF94s give both good bond lengths and bond angles. The bond angle predictions from HF 3-21G(*) are of modest quality (Figure 6). A few of the force fields do not give very accurate bond length and bond angle predictions. For both bond lengths (Figure 3) and bond angles (Figure 6), the Sybyl, Tripos, and OPLS-AA force fields are consistently among the poorest performers.

A careful judgment has to be made in regard to treating the conformational variables. There is only meager experimental information on the ATMO conformational preferences available which we will discuss later. In the case of bond lengths and bond angles, we had experimental means with which to compare computed results. However, we have no experimental torsional potential energy curves. Often, Hartree−Fock theory gives a very reasonable potential energy curve for simple cases such as the internal rotation of ethane. A priori we do not know if the same will be true for ATMO. Users of computational chemistry methods commonly assume that going to a higher level of theory will give better results. However, in the case of the optimized bond lengths and bond angles of ATMO, we have seen that this is not a safe assumption. Hence rather than assume the theoretically predicted conformational energy curves from any one method are best, we have chosen to determine consensus potential energy curves to use as references for comparison of the 27 methods. There is precedent for using a consensus to compare methods. For example, in the early days of computational chemistry a consensus prediction was sometimes used when the only methods available were approximate molecular orbital methods of limited accuracy. More recently, a consensus was sometimes used to compare methods for docking predictions.[54,55] Nevertheless, we realize that use of consensus curves as the reference torsional potentials is arbitrary.

It is worth recalling that our goal is to represent a potential energy curve corresponding to rotation about a single bond by a set of numbers that can be used as components of a vector. The vector will be used in the cluster analysis. In effect, we need to represent a continuous curve by a finite set of discrete numbers, in effect, creating a "fingerprint" to describe the conformational characteristics of molecule. A first possible approach, which we prefer and use, is to scan the torsion uniformly across 360° using moderately sized dihedral angle increments. This will identify energy valleys and rotational barriers, which are the major determinants of conformation. A second possibility would be to scan the rotational potential in very small increments just close to the energy minima. However, our purpose is not to try to reproduce only the harmonic part near the bottom of the potential energy curves. Moreover, with this alternative, one would still need to do the broad scan to locate the valleys. A third possibility would be to rate the computational chemistry methods in terms of how closely they reproduce crystallographic torsional angles when the starting conformations are the solid-state ones. However, this third approach could be complicated by the influence of crystal packing forces. Use of this latter alternative approach is the subject of ongoing research.

Because the ability to predict the modality of a potential curve (1-fold, 2-fold, etc.) correctly is of major importance to us, we have chosen the first approach described above. For scanning the potential energy curve, we use an increment of 30° which is a reasonable compromise between providing enough detail about the potential energy surfaces and completing all the single-point calculations in a reasonable time span. As mentioned, we study three rotatable bonds in our ATMO model structure. Relative energies within the each method are obtained by setting the lowest energy to zero in the data set for each rotatable bond. As expected, the calculated energies are extremely high at sterically prohibited conformations. In fact, for a few such conformations the energy could not be computed, i.e., the calculations fail because interatomic contacts were too close (these cases are marked as ND for no data in Tables S6-S8). The exact values of these very high energies are not significant, so all high energies were truncated. Initially we tried truncation at 100 kcal/mol relative energy. This seemed to work satisfactorily, although it can be argued that it gives too much weight to the high energy tails of the potential energy curves. Hence a value lower than 100 may be in order, but not so low that the tops of many minor peaks would be cut off. In the interest of retaining as much of the computed energy data as possible, we settled on truncating at 25 kcal/mol. After converting the resulting energies to eV, an average of the relative single-point energies at each 30° increment over all 27 methods was computed. The average is referred as the consensus in this paper. As alluded to earlier, we want the consensus to reflect both the conformational detail in the valleys and the foldedness of the potential energy curves. Results are presented in the potential energy curves in Figure 8. Included in Figure 8 are the wheel diagrams with the dihedral angles observed in the crystal structures of Table 1.

It can be seen in the wheel diagram of Figure 8(a) that the experimental values of the C6−O1−N3−C4 dihedral angle cluster near 180°. Thus, the methyl group of the methoxime is trans. Delocalization would favor a coplanar methoxime, but at 0° the methyl group would bump into H2 (Figure 2). The energy curves of Figure 8(a) are completely consistent with this. The only minimum is near 180°; a high barrier to internal rotation occurs near 0°.

The wheel diagram in Figure 8(b) shows clusters near 0° and 180° for N1−C2−C4−N3. This observation corresponds to the thiazole ring flipping 180° about the C2−C4 bond between roughly coplanar syn and anti arrangements. Computationally, the only consistently calculated energy minimum is near 180°. This is the dihedral angle where 9 of the 12 crystal structures cluster. For rotation around the C2−C4 bond, none of the methods have minimum between −30 and +30°, but Tripos has a very low relative energy of only 0.5 kcal/mol at −30°, and MMFF94 (Spartan) gives low relative energies of 1.2 kcal/mol at 30° and 1.9 kcal/mol at 0°. Thus, these force field implementations suggest it would be relatively easy to achieve a conformation near 0° such as might be induced by crystal packing forces.

We see what appears to be four clusters in the wheel diagram of Figure 8(c), roughly at 0°, 90°, 180°, and 270° for rotation around C4−C5. Thus, none of the three rotatable bonds in ATMO is 3-fold. The dominant cluster (7 cases) corresponds to a dihedral angle near 270°. A few of the methods are consistent with this observation. For instance, the semiempirical methods (AM1, PM3, MNDO) predict a minimum at 270°, and all MMFF implementations show very low energies at this dihedral angle, e.g., MMFF94 in MacroModel gives a relative energy of 0.1 kcal/mol at 270°. Compared to the curves in Figure 8(a),(b), the consensus in Figure 8(c) has few high peaks. This result suggests that rotation around the C4−C5 bond is easier than at the O1−N3 and C2−C4 bonds. In harmony with this, PM3, MMFF94 (Spartan), Sybyl (Spartan), DFT 6-31G*, DFT 6-31G**, DFT 6-311G*, DFT 6-31+G*, DFT 6-311+G**, AMBER*, OPLS/A, MMFF94s (MMod), Tripos, MMFF94 (SYBYL),

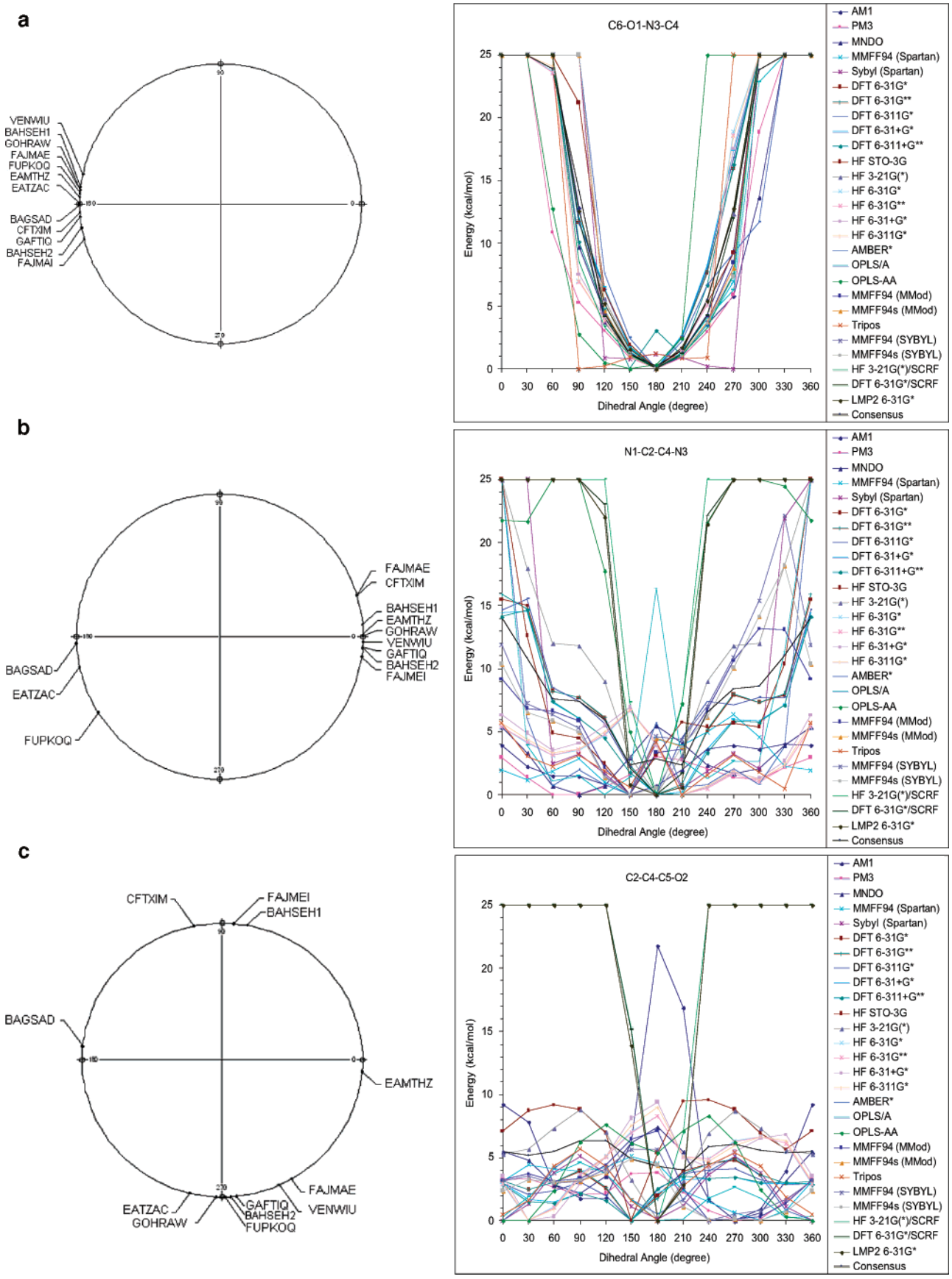

Figure 8. (a) Experimental dihedral angle values (in degrees) and computed potential energy curves (in kcal/mol) (from single-point energy calculations) for C6−O1−N3−C4. (b) Experimental dihedral angle values (in degrees) and computed potential energy curves (in kcal/mol) (from single-point energy calculations) for N1−C2−C4−N3. (c) Experimental dihedral angle values (in degrees) and computed potential energy curves (in kcal/mol) (from single-point energy calculations) for C2−C4−C5−O2. For each of the three rotatable bonds, the energies were truncated at 25 kcal/mol as explained in the text.

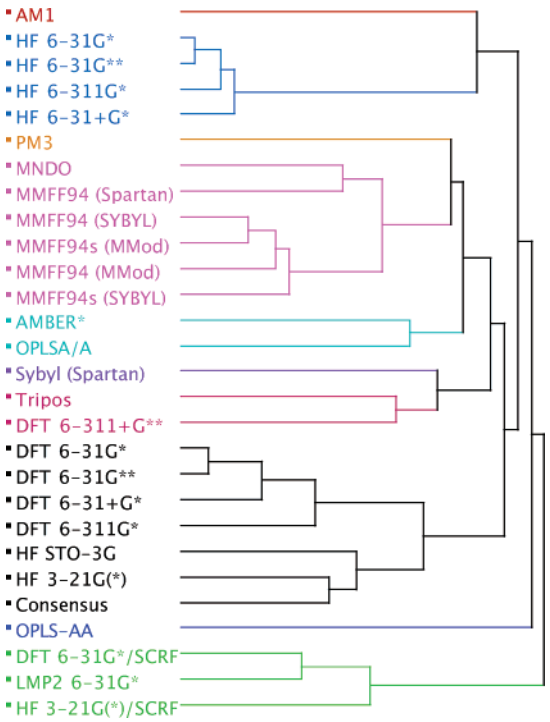

Figure 9. Dendrogram obtained using JMP for analysis of dihedral angles. The plot is based on single-point energies truncated at 25 kcal/mol, for dihedral angles in the range 0−360°, for all three rotatable bonds taken together.

and MMFF94s (SYBYL) all predict potential energy curves that do not rise above 6 kcal/mol in the entire range 0−360°. And MMFF94 (MMod) gives only two points above 6: 6.48 and 7.12 kcal/mol. We may conclude that the *N*-methylacetyl moiety has more flexibility than the other side chains, and perhaps it adopts several different conformations influenced by the crystalline state environment. The limited amount of X-ray data precludes us from speaking more definitively about the shape of the potential energy curves for internal rotation.

It is evident from the data in Figure 8 that there is considerable variability in the details in the potential energy curves predicted by the 27 methods. This finding supports our decision to not rely on any single method as producing "correct" potential energy curves. Hence we will proceed to use the consensus curves as references for the cluster analysis. As seen in the dendrogram of Figure 9, the quantum mechanical HF 3-21G(*) method correlates best with the consensus potential energy curves for the three rotatable bonds. The HF STO-3G and the DFT results are next closest to the consensus. MMFF94 in MacroModel is the best force field for giving what appear to be reasonable single-point energies of the ATMO conformers. The OPLS/A force field appears to be better than OPLS-AA in terms of rotational potentials. Two of the semiempirical molecular orbital methods (MNDO and PM3) are almost as good as the MMFF implementations. Ab initio optimizations with large basis sets do not give results very close to the consensus. We also see in Figure 9 that adding the effect of solvation or going to a post-Hartree−Fock method gives worse agreement with the consensus.

As an aside, we mention that with the original 100 kcal/mol energy cutoff, instead of 25 kcal/mol, HF 3-21G(*) and HF STO-3G are closest to the consensus, and the methods

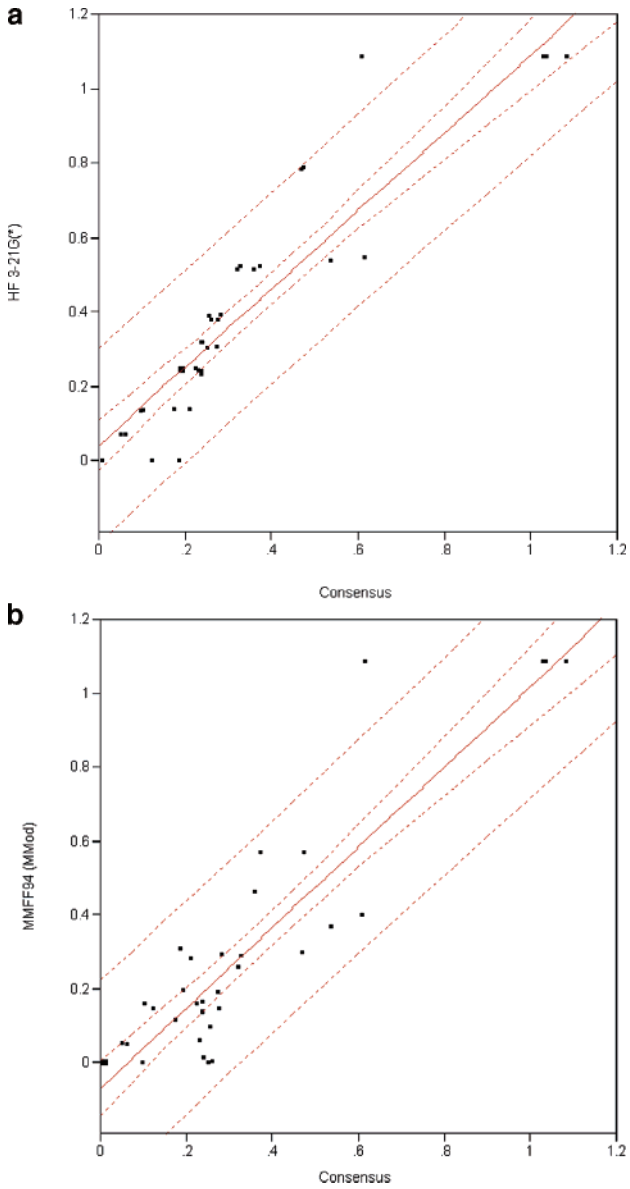

Figure 10. (a) Bivariate fit of HF 3-21G(*) vs consensus of potential energy curves for internal rotation. The outlier with respect to the upper and lower 95% confidence curves for individual data points is 0° for the N1−C2−C4−N3 dihedral angle. Regression equation: HF 3-21G(*) = 0.0428 (±0.0330) + 1.0490 (±0.0675) consensus; n = 36; r^2 = 0. 8765; r_{adj}^2 = 0.8729; RMS error = 0.1246; t ratios = 1.30, 15.54; p < 0.0001. (b) Bivariate fit of MMFF94 (in MacroModel) vs consensus of potential energy curves for internal rotation. The outlier with respect to the upper and lower 95% confidence curves for individual data points is 90° for the C6−O1−N3−C4 dihedral angle. Regression equation: MMFF94 (MMod) = −0.0685 (±0.0368) + 1.0881 (±0.0753) consensus; n = 36; r^2 = 0. 8599; r_{adj}^2 = 0. 8557; RMS error = 0. 1391; t ratios = −1.86, 14.44; p < 0.0001.

including solvation and the post-Hartree−Fock method are farthest from the consensus. Thus the choice of cutoff value did not affect the outcome regarding the best and worst method.

Regression analyses of the HF 3-21G(*) and MMFF94 results versus the consensus are presented in Figure 10. These two methods are in fair accord with the consensus. Note that the amount of variance explained by the correlations drops a little below 90% but is still quite good compared to what is typically encountered in studies of quantitative structure−activity relationships.

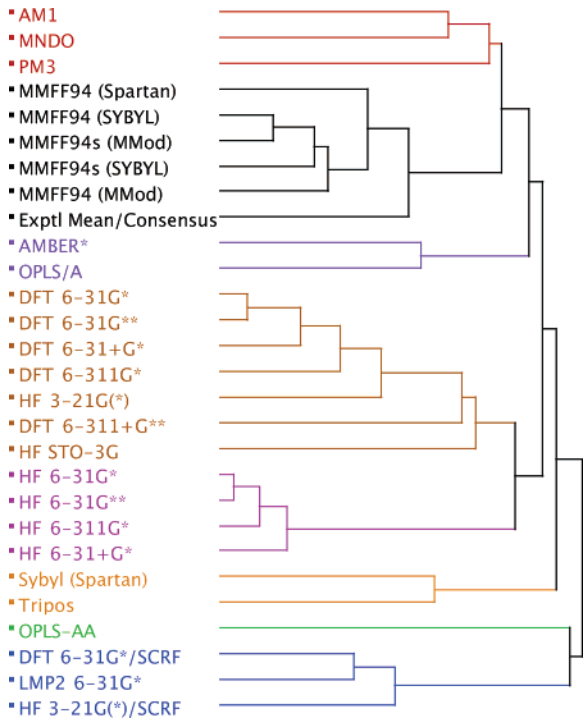

Figure 11. Dendrogram obtained using JMP for analysis of the combination of experimental bond length, bond angle, and consensus of the single-point conformational energy data.

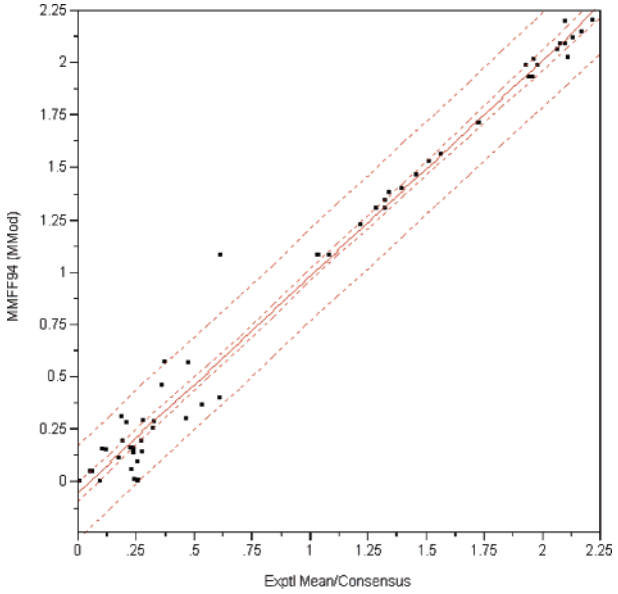

Figure 12. Bivariate fit of MMFF94 (in MacroModel) vs the experimental mean and consensus data. The singular outlier with respect to the upper and lower 95% confidence curves for individual data points is for the C6−O1−N3−C4 dihedral angle at 90°. Three data points fall near the 95% confidence limits: the N1−C2−C4−N3 dihedral angle at 300° and the C2−C4−C5−O2 dihedral angle at 240° and 270°. Regression equation: MMFF94 (MMod) = −0.0476 (±0.0225) + 1.030 (±0.0186) exptl mean/consensus, $n = 61$; $r^2 = 0.9811$; $r_{adj}^2 = 0.9807$; RMS error = 0.1093; t ratios = −2.11, 55.27; $p < 0.0001$.

When bond lengths, bond angles, and the consensus of the single-point energies are considered together (Figure 11), cluster analysis shows that molecular mechanics by the various implementations of the MMFF94 force field give the best overall agreement with all the reference data. The Hartree−Fock and density functional theory results are roughly comparable to each other in quality. Adding the effect of solvation did not improve agreement with solid-state data. The success of MMFF94 in treating the ATMO structure is also shown by the bivariate fit in Figure 12. The correlation coefficient is good, and the t ratios for the two terms of the regression equation are very good (significantly greater than 2 in magnitude). Only one of the 61 data points is clearly outside the 95% confidence curves for individual data points.

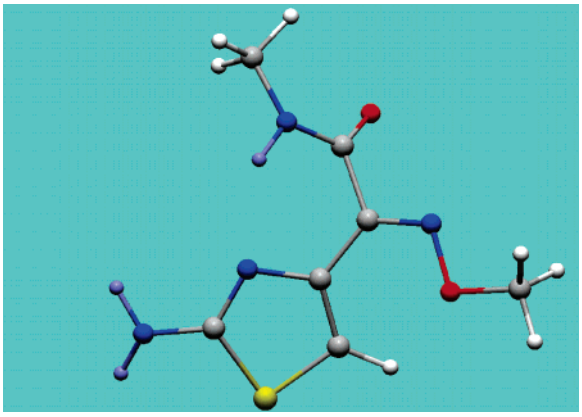

Figure 13. Final optimized geometry obtained using MMFF94 in MacroModel. The ball-and-stick depiction is from MacroModel.

The optimized geometry predicted by MMFF94 is shown in Figure 13. Except for the methyl hydrogens, the structure is roughly coplanar.

CONCLUSIONS

In this work, comparison of a wide range of currently popular quantum mechanical and molecular mechanics methods reveals that the most reliable approach for reproducing the reference data of the diverse bond types in the ATMO substructure is molecular mechanics. Whereas the finding that molecular mechanics is better than most of the quantum methods investigated may be surprising to some theoreticians who have not examined force field methods, the finding portends well for more efficient molecular modeling. Molecular mechanics enjoys the important advantages of being much less demanding of computer resources and, in the case of the pharmaceutically related molecule examined here, being more reliable or just as reliable.

Our findings appear to call into question the common practice of "validating" a theoretical prediction by repeating the calculation at a higher or newer level of theory. Instead, we recommend a greater diversity of methods (both quantum mechanical and force field) be used in attempts to validate a prediction.

This work also illustrates that when the goal is to generate accurate molecular structures on the computer careful consideration has to be given to the selection of the theoretical method. This reminder is apropos in this age of high throughput virtual screening. The computational chemistry method must be selected by balancing desires for accuracy and generality as well as speed. For instance, a recent paper using HIV-1 reverse transcriptase inhibitors as test cases for evaluating a new docking protocol sensibly invoked MMFF94 as their force field of choice.[56]

For our particular case of ATMO, solvation did not play a major role. However, for other pharmaceutically interesting molecules, solvation will have a more significant effect on conformation, although not necessarily on bond lengths and bond angles.

Because molecular mechanics by the MMFF94 force field is successful at closely reproducing the crystallographic data for the aminothiazole methoxime substructure, we conclude that it is a reliable method to pursue further molecular modeling analysis of ATMO-containing materials.

A computational chemistry study may sometimes reveal the need for further experimental work. In our case, we challenge crystallographers to solve more ATMO-containing structures at better resolution. Such new data would be helpful in evaluating and refining our analysis.

Finally, we recommend cluster analysis for comparing the performance of computational chemistry methods on other molecules. This cheminformatics approach has the advantages of being fast, presenting its quantitative findings in easy-to-assimilate graphical format, and showing an overview of all the data compared.

ACKNOWLEDGMENT

We thank Dr. Kelsey Forsythe for his diligent management of the computational chemistry programs and workstations. We are grateful to Professor Stephanie Sen for her interest. We also acknowledge Dr. Dale Braden at the Schrödinger technical support desk for resolving software issues.

Supporting Information Available: Tables of experimental bond lengths, bond angles, and dihedral angles of the ATMO-containing compounds, theoretically determined bond lengths and bond angles, and single-point energies of conformers of the ATMO-containing test model. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) O'Neil, M. J.; Smith, A.; Heckelman, P. E.; Budavari, S. *Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*; Merck & Co. Inc.: Whitehouse Station, NJ, 2001.

(2) Webber, J. A.; Wheeler, W. J. Antimicrobial and Pharmacokinetic Properties of Newer Penicillins and Cephalosporins. In *Chemistry and Biology of Beta-Lactam Antibiotics;* Morin, R. B., Gorman, M., Eds.; Academic Press: New York, 1982; Vol. 1, pp 371−436.

(3) Dürckheimer, W.; Adam, F.; Fischer, G.; Kirrstetter, R. Synthesis and Biological Properties of Newer Cephem Antibiotics. In *Frontiers of Antibiotic Research;* Umezawa, H., Ed.; Academic Press: New York, 1987; pp 161−192.

(4) Baertschi, S. W.; Cantrell, A. S.; Kuhfeld, M. T.; Lorenz, L. J.; Boyd, D. B.; Jaskunas, S. R. Inhibition of Human Immunodeficiency Virus Type 1 Reverse Transcriptase by Degradation Products of Ceftazidime. *Antiviral Chem. Chemother*. **1997**, *8*, 353−362.

(5) *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P., Schaefer H. F. III, Eds.; Wiley: Chichester, UK, 1998; Vols. 1−5.

(6) Jensen, F. *Introduction to Computational Chemistry*; Wiley: Chichester, UK, 1999.

(7) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; Wiley: Hoboken, NJ, 2002.

(8) Pettersson, I.; Liljefors, T. Molecular Mechanics Calculated Conformational Energies of Organic Molecules: A Comparison of Force Fields. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1996; Vol. 9, pp 167−189.

(9) Liljefors, T.; Gundertofte, K.; Norrby, P.; Petterson, I. Molecular Mechanics and Comparison of Force Fields. In *Computational Medicinal Chemistry for Drug Discovery*; Bultinck, P., Winter, H. D., Langenaeker, W., Tolleraere, J. P., Eds.; Marcel Dekker: New York, 2004; Vol. 1, pp 1−28.

(10) Alagona, G.; Ghio, C.; Monti, S. B3LYP/6-31G* Conformational Landscape in Vacuo of Some Pterocarpan Stereoisomers with Biological Activity. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2849−2857.

(11) Blatchly, R. A.; Tew, G. N. Theoretical Study of Helix Formation in Substituted Phenylene Ethynylene Oligomers. *J. Org. Chem.* **2003**, *68*, 8780−8785.

(12) Downs, G. M.; Barnard, J. M. Clustering Methods and Their Uses in Computational Chemistry. In *Reviews in Computational Chemistry;* Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: Hoboken, NJ, 2002; Vol. 18, pp 1−34.

(13) Boyd, D. B. On the Rhodanines and Their Presence in Biologically Active Ligands. *J. Mol. Struct.* **1997**, *401*, 227−234.

(14) ConQuest 1.6 User Guide. Cambridge Crystallographic Data Centre, Cambridge, CB2 1EZ, UK. (http://www.ccdc.cam.ac.uk/products/csd_system/conquest/).

(15) Laurent, G.; Durant, F. Structure of 2-(2-Amino-4-thiazolyl)-2(*E*)-(methoxy)iminoacetic Acid C6H7N3O3S. *Cryst. Struct. Commun.* **1981**, *10*, 1007−1014.

(16) Laurent, G.; Durant, F. Structure of 2-(2-Amino-4-thiazolyl)-2(*Z*)-(methoxy)iminoacetic Acid C6H7N3O3S.3H2O. *Cryst. Struct. Commun.* **1981**, *10*, 1015−1023.

(17) Laurent, G.; Evrard, G.; Durant, F. Preliminary Results on the X-ray Structure Analysis of Oxyimino Cephalosporins. *Eur. J. Med. Chem.* **1982**, *17*, 281−284.

(18) Laurent, G.; Parmentier, C.; Durant, F.; Evrard, G. Structure of 2-(2-Amino-4-thiazolyl)-2(*Z*)-(methoxy)imino Ethyl Acetate, The Side Chain of Cefotaxime, A Highly Active Cephalosporin C8H11N3O3S. *Cryst. Struct. Commun.* **1981**, *10*, 281−287.

(19) Laurent, G.; Parmentier, C.; Durant, F.; Evrard, G. Ethyl 2-Amino-alpha-(*E*-methoxyimino)-4-thiazoleacetate Hydrobromide. *Acta Crystallogr., Sect. B* **1981,** *37*, 974−976.

(20) Laurent, G.; Durant, F.; Evrard, G. Ethyl 2-Amino-alpha-(*E*-methoxyimino)-4-thiazoleacetate. *Acta Crystallogr., Sect. B* **1981**, *B37*, 972−974.

(21) Miyamae, A.; Koda, S.; Morimoto, Y. The Crystal and Molecular Structures of Ceftizoxime and Ceftizoxime Monohydrochloride Monohydrate. *Chem., Pharm. Bull.* **1986**, *34*, 3539−3548.

(22) Singh, J.; Fox, R.; Wong, M.; Kissick, T. P.; Moniot, J. L.; Gougoutas, J. Z.; Malley, M. F.; Kocy, O. The Structures of Alkoxycarbonyl, Acyl, and Sulfonate Derivatives of 1-Hydroxybenzotriazole: Nitrogen- vs Oxygen-Substitution. *J. Org. Chem.* **1988**, *53*, 205−208.

(23) Laurent, G.; Norberg, B.; Durant, F. Crystal Structure of 2-(2-Amino-4-thiazolyl-5-chloro)-2-*Z*-Methoxyimino-*N*-methylacetamide, C7O2N4SClH9. *Bull. Soc. Chim. Belg.* **1987**, *96*, 747−749.

(24) Laurent, G.; Norberg, B.; Durant, F. Crystal Structure of 2-(2-Amino-4-thiazolyl)-2(*Z*)-(methoxy)imino Methylacetamide Hydrochloride Monohydrate, An Analogue to the C7 Side Chain of Cefotaxime. *Bull. Soc. Chim. Belg.* **1988**, *97*, 325−327.

(25) Yoshida, A.; Moroi, R.; Arimoto, M.; Furukawa, M. Crystal Structure of 7-[(*Z*)-2-(2-Aminothiazol-4-yl)-2-(imidazol-4-ylmethoxyimino)acetamido]-3-[(1-pyridinio)methyl]-3-cephem-4-carboxylate Diperchlorate Dihydrate. *Anal. Sci.* **1989**, *5*, 785−786.

(26) Deguchi, S.; Fujioka, M.; Okamoto, Y.; Yasuda, T.; Nakamura, N.; Yamaguchi, K.; Suzuki, S. Structural Studies on an Iron(III) Complex Containing (*Z*)-2-(2-Aminothiazol-4-yl)-N−(2-hydroxyethyl)-2-(hydroxyimino)acetamide, A Model Compound for a Cephalosporin Antibiotic Cefdinir. *J. Chem. Soc., Dalton Trans.: Inorg. Chem.* **1996**, *9*, 1967−1971.

(27) Miyamae, A.; Koda, S.; Morimoto, Y. *Acta Crystallogr., Sect. A* **1984**, *40*, C76.

(28) JMP 5.1 Tutorial, SAS Institute Inc, Cary, NC 27513. (http://www.jmp.com/)

(29) Autodesk, San Rafael, CA 94903. (http://usa.autodesk.com/)

(30) Spartan '04 Windows Tutorial and User's Guide. Wavefunction Inc.: Irvine, CA 92612. (http://www.wavefun.com/software/spartan04_win/win_manual/main.html).

(31) SYBYL Version 6.8 Manual. Tripos Inc.: St. Louis, MO 63144. (http://www.tripos.com/sciTech/inSilicoDisc/emmt.html).

(32) MacroModel Version 7.0 User's Manual. Schrödinger Inc.: Portland, OR 97201. (http://www.schrodinger.com/Products/macromodel.html).

(33) Jaguar 5.5 User's Manual. Schrödinger Inc.: Portland, OR 97201. (http://www.schrodinger.com/Products/jaguar.html).

(34) Stewart, J. J. P. Semiempirical Molecular Orbital Methods. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1990; Vol. 1, pp 45−81.

(35) Zerner, M. C. Semiempirical Molecular Orbital Methods. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1991; Vol. 2, pp 313−365.

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

(37) Feller, D.; Davidson, E. R. Basis Set for Ab Initio Molecular Orbital Calculations and Intermolecular Interactions. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1990; Vol. 1, pp 1−43.

(38) Young, D. *Computational Chemistry A Practical Guide for Applying Techniques to Real World Problems*; Wiley: New York, 2001.
(39) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(40) Halgren, T. A. The Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827−7843.
(41) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(42) Halgren, T. A. Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(43) Halgren, T. A. Merck Molecular Force Field. III. Molecular Geometrics and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586.
(44) Halgren, T. A.; Nachbar, R. B. Merck Molecular Force Field. IV. Conformational Energies and Geometries. *J. Comput. Chem*. **1996**, *17*, 587−615.
(45) Halgren, T. A. Merck Molecular Force Field. V. Extension of MMFF94 using Experimental Data, Additional Computational Data and Empirical Rules. *J. Comput. Chem.* **1996**, *17*, 616−641.
(46) Clark, M.; Cramer, R. D. III; van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989,** *10*, 982−1012.
(47) Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J. Comput. Chem.* **1999**, *20*, 720−729.
(48) McDonald, D. Q.; Still, W. C. AMBER* Torsional Parameters for the Peptide Backbone. *Tetrahedron Lett.* **1992**, *33*, 7743−7746.
(49) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(50) Jorgensen, W. L.; Maxwell D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(51) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All-Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem*. **1986**, *7*, 230−252.
(52) Pulay, P.; Saeboe, S. Orbital-Invariant Formulation and Second-Order Gradient Evaluation in Moeller−Plesset Perturbation Theory. *Theor. Chim. Acta* **1986**, *69*, 357−368.
(53) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A. III; Honig, B. Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. *J. Am. Chem. Soc*. **1994**, *116*, 11875−11882.
(54) Muegge, I.; Rarey, M. Small Molecule Docking and Scoring. In *Reviews in Computational Chemistry;* Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2001; Vol. 17, pp 1−60.
(55) Boehm, H.-J.; Stahl, M. The Use of Scoring Functions in Drug Discovery Applications. In *Reviews in Computational Chemistry;* Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: Hoboken, NJ, 2002; Vol. 18, pp 41−87.
(56) de Jonge, M. R.; Koymans, L. M. H.; Vinkers, H. M.; Daeyaert, F. F. D.; Heeres, J.; Lewi, P. J.; Janssen, P. A. J. Structure Based Activity Prediction of HIV-1 Reverse Transcriptase Inhibitors. *J. Med. Chem*. **2005**, *48*, 2176−2183.

CI049671X