

Molecular Modeling of Geometries, Charge Distributions, and Binding Energies of Small, Druglike Molecules Containing Nitrogen Heterocycles and Exocyclic Amino Groups in the Gas Phase and in Aqueous Solution

Brian R. White,[†] Carston R. Wagner,^{†,‡} Donald G. Truhlar,[‡] and Elizabeth A. Amin^{*,†}

Department of Medicinal Chemistry, College of Pharmacy, and Department of Chemistry and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455

Received March 5, 2008

Abstract: We have tested a variety of approximate methods for modeling 30 systems containing mixtures of nitrogen heterocycles and exocyclic amines, each of which is studied with up to 31 methods in one or two phases (gaseous and aqueous). Fifteen of the systems are protonated, and fifteen are not. We consider a data set consisting of geometric parameters, partial atomic charges, and water binding energies for the methotrexate fragments 2-(aminomethyl)pyrazine and 2,4-diaminopyrimidine, as well as their cationic forms 1H-2-(aminomethyl)pyrazine and 1H-2,4-diaminopyrimidine. We first evaluated the suitability of several density functionals with the 6–31+G(d,p) basis set to serve as a benchmark by comparing calculated molecular geometries to results obtained from coupled-cluster [CCSD/6–31+G(d,p)] wave function theory (WFT). We found that the M05-2X density functional can be used to obtain reliable geometries for our data set. To accurately model partial charges in our molecules, we elected to use the well-validated charge model 4 (CM4). In the process of establishing benchmark values, we consider gas-phase coupled cluster and density functional theory (DFT) calculations, followed by aqueous-phase DFT calculations, where the effect of solvent is treated by the SM6 quantum mechanical implicit solvation model. The resulting benchmarks were used to test several widely available and economical semiempirical molecular orbital (SE-MO) methods and molecular mechanical (MM) force fields for their ability to accurately predict the partial charges, binding energies to a water molecule, and molecular geometries of representative fragments of methotrexate in the gaseous and aqueous phases, where effects of water were simulated by the SM5.4 and SM5.42 quantum mechanical implicit solvation models for SE-MO and explicit solvation was used for MM. In addition, we substituted CM4 charges into the MM force fields tested to observe the effect of improved charge assignment on geometric and energetic modeling. The most accurate MM force fields (with or without the CM4 charges substituted) were validated against gas-phase and aqueous-phase geometries and charge distributions of a larger set of 16 druglike ligands, both neutral and cationic. This process showed that the Merck Molecular Force Field (MMFF94) with or without CM4 charges substituted, is, on average, the most accurate force field for geometries of molecules containing nitrogen heterocycles and exocyclic amino groups, both protonated and unprotonated. This force field was then applied to the complete methotrexate molecule, in an effort to systematically explore its accuracy for trends in geometries and charge distributions. The most accurate force fields for the binding energies of nitrogen heterocycles to a water molecule are OPLS2005 and AMBER.

1. Introduction

Continuing advances in molecular modeling and computational chemistry have greatly facilitated the structure-based

design of small-molecule inhibitors of proteins.^{1–15} Although molecular mechanics (MM) force fields^{16–20} can model protein structure, they often lack parameters that accurately represent the heteroatomic groups present in pharmaceuticals.^{21–23} Density functional theory²⁴ (DFT) and wave function theory (WFT)²⁵ do not require new parameters for each type of atom; however, current technology still limits the calculations to smaller molecules and exploratory studies

* To whom correspondence should be addressed. E-mail: eamin@umn.edu.

[†] Department of Medicinal Chemistry, College of Pharmacy.

[‡] Department of Chemistry and Minnesota Supercomputing Institute.

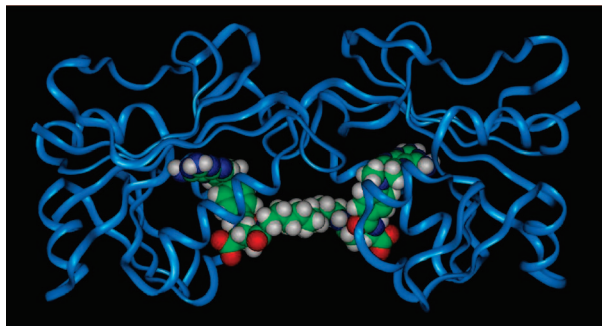


Figure 1. DHFR₂MTX₂ chemically induced dimer.

on larger systems. Two viable approaches for simulating a protein bound to a druglike inhibitor are to obtain MM parameters for force fields that yield accurate molecular geometries and partial charges or to find a suitable level of combined QM/MM theory^{26–28} in which a critical or active region of the system is treated by quantum mechanics (QM) and the surrounding areas by MM. An economical QM level for such calculations would be semiempirical molecular orbital^{29–34} (SE-MO) theory. For small enough QM regions or short simulations, one can also use more reliable QM methods such as DFT. Reliable WFT calculations are, however, affordable only for the smallest systems.

A recent application of molecular modeling is the prediction of mutation effects on protein–protein interactions.^{35–42} Protein multimer stability can be modified through the introduction of interfacial residue mutations, and it would be valuable to be able to predict the relative change in stability of a mutated protein multimer compared to the wild-type species. Such calculations would aid in understanding the functional evolution of proteins, as well as the development and control of stable, self-assembled protein structures, with applications ranging from nanoscale multiprotein constructs to drug delivery. With the advent of chemically induced dimerization (CID), our laboratory has demonstrated the ability to create self-assembled *E. coli* dihydrofolate reductase (DHFR) dimers from naturally existing DHFR monomers using a bivalent methotrexate dimerizer (MTX₂, complex–DHFR₂MTX₂) (Figure 1).⁴³ While the nature and effects of linker length have been examined previously, we are still in the early stages of in vitro and in silico observation of interfacial mutation effects on the dimer. We have found that the introduction of complementary interfacial mutations putatively leads to a stabilized DHFR heterodimer, which allows for a level of control over the assembly of such constructs.

One complicating issue present in our system, as well as other biological systems, is the protonation state of the ligand in solution and in complex with the protein. While the DHFR inhibitor MTX is unprotonated in solution, it is protonated on N1 (Figure 2) when bound to DHFR.^{44,45} This raises the question of whether it is appropriate to use a single set of MM parameters to describe MTX both in solution and bound to the enzyme. To assist the in silico prediction of mutation effects on dimer stability, we have undertaken a study to develop a set of MM parameters that can accurately model the DHFR₂MTX₂ complex. To accomplish this goal, we will first try to establish an accurate method to model the single

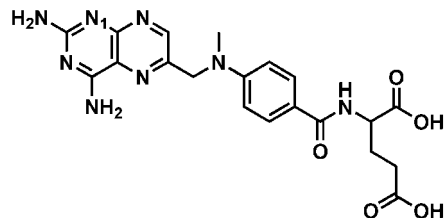


Figure 2. Structure of methotrexate.

substrate, MTX, and then try to extend this method to the DHFR₂MTX₂ complex. During this process, we have tested a large variety of methods on a set of druglike molecules containing nitrogen heterocycles and exocyclic amino groups, and the results of these tests are presented in the present article because they should be of general interest for a variety of potential applications.

A critical issue in simulating systems with nitrogen heterocycles is modeling the charge distributions. Because partial charge distributions are not experimental observables, we will rely on theory to establish reasonable values. For this purpose, we first require accurate geometries, and we begin by establishing benchmark values using high-level WFT and DFT calculations on the MTX fragments 2-(aminomethyl)pyrazine (2-AMP), 1*H*-2-(aminomethyl)pyrazine (1*H*-2-AMP), 2,4-diaminopyrimidine (2,4-DAP), and 1*H*-2,4-diaminopyrimidine (1*H*-2,4-DAP). These fragments were chosen because of the role of the pteridine moiety in MTX binding to the DHFR active site. We then use DFT with class IV charges⁴⁶ to establish benchmark partial atomic charges. Coupled cluster theory^{47,48} with single and double excitations (CCSD) and the M05-2X⁴⁹ density functional with the 6–31+G(d,p)⁵⁰ basis set are used for geometries, and charge model 4⁵¹ (CM4) is used for partial atomic charges. The performance of widely available SE-MO and MM parametrizations is then surveyed for these four fragments to find the parametrized model that most accurately predicts the geometries, binding energies to water, and charge distribution of the unprotonated and protonated states of 2-AMP and 2,4-DAP. In addition, we consider the selected MM methods when CM4 charges are substituted for the force field's default charges in an effort to observe if increased accuracy in partial charge distribution leads to increased performance in geometric and energetic modeling. The most accurate methods are then used to calculate partial charges and geometries for a series of pharmacophorically similar molecules containing nitrogen heterocycles and exocyclic amines, and these results are compared to DFT and CM4 benchmarks to explore the validity of our chosen MM parameters more broadly.

2. Methods and Software

2.1. Computational Methods. For geometries, we consider three categories of QM theory plus MM. The QM categories are WFT, DFT, and SE-MO. For partial charges, we consider Mulliken population analysis,⁵² MM^{53–57} and the CM1,⁴⁶ CM2,⁵⁸ CM3,⁵⁹ and CM4⁵¹ charge models. DFT calculations in the aqueous phase use the implicit solvation model SM6,⁵¹ while solvation in SE-MO methods^{30–32} was included by using the implicit SM5.4⁶⁰ or SM5.42⁶¹ solvation

models. For aqueous-phase MM calculations, we employed explicit solvation using the respective programs' soak algorithms in conjunction with periodic boundary conditions (minimum cell size of $15 \times 15 \times 15$ Å) to eliminate solvent–vacuum interfaces.

WFT calculations were carried out by CCSD with the 6–31+G(d,p)⁵⁰ basis set. These calculations were carried out with the *Gaussian03* computer program (Gaussian, Inc.).⁶² The CCSD method was chosen over the popular Møller–Plesset second-order perturbation theory⁶³ as CCSD (or the closely related QCISD⁶⁴) has been shown to yield more accurate geometries.⁶⁵

The DFT methods examined are B3LYP,^{66,67} mPW1PW^{68,69} (which is also called mPW1PW91, mPW0, and MPW25), MPWB1K,^{67,69,70} MPW1KCIS,⁷¹ M06-L,⁷² and M05-2X⁴⁹ with the 6–31+G(d,p)⁵⁰ basis set. Calculations were performed using a locally modified version of the *Gaussian03* program incorporating the MN-GSM 6.0⁷³ and MN-GFM 2.0.1⁷⁴ solvation and DFT modules. To select a density functional to generate benchmark values, gas-phase calculations were carried out, and the resulting geometries were evaluated relative to CCSD. The best functional was subsequently used to perform calculations in the aqueous phase using the SM6⁵¹ solvation model for implicit solvation.

We have tested charge model 4 (CM4) partial charge assignments based on gas-phase and SM6 DFT calculations. CM4 charges, the fourth generation of class IV⁴⁶ charges, have a distinct advantage over the class II^{52,75–77} and III^{78–80} charges used in *Gaussian03*. Whereas the reliability of class III charges depends on the wave function and basis set used, class IV charges represent an extrapolation to full configuration interaction with a complete basis set.^{46,51} Furthermore, class III charges are unstable with respect to buried charges,^{81–84} while class IV charges provide a reliable method for obtaining buried charges. CM4 charges, in particular, have been parametrized against a large training set (398 molecules) and are well suited for modeling aliphatic functional groups, which makes them more suitable for modeling hydrophobic effects, a primary factor in protein–protein interactions.

SE-MO methods examined in the current study include AM1,³⁰ PM3,^{31,32} and PDDG/PM3.⁸⁵ Calculations were performed for AM1 and PM3 using AMSOL 7.1⁸⁶ (a derivative of AMPAC 2.1) with Mulliken, CM1, CM2, or CM3 charges obtained from gas-phase calculations. For aqueous-phase AM1 and PM3 calculations, AMSOL 7.1 was used to obtain Mulliken, CM1, or CM2 charges within the SM5.4 (for Mulliken and CM1) or the SM5.42 (for CM2) solvation models. GAMESSPLUS⁸⁷ was used to obtain a second set of CM3 charges in the gas phase in an effort to test consistency in charge assignment across software. The notation used in this article for AM1 calculations with differing partial charge assignments is AM1, AM1-CM1, AM1-CM2, and AM1-CM3 for AM1 calculations with Mulliken, CM1, CM2, and CM3 charges, respectively. The notation for the PM3 calculations is analogous to that for AM1. It is important to note that because they are post-self consistent field (SCF) analysis tools, CMx or Mulliken charges of gas-phase wave functions do not alter an

optimized molecule's geometry. Slight differences may be attributed to variances in the convergences of the SCF and geometry optimizations. In solution, each SMx model uses a particular choice of charge model; this choice, along with all the other SMx parameters, does affect the molecules' geometry. PDDG/PM3 gas-phase optimizations were performed using MOPAC 5.011mn,⁸⁸ and Mulliken partial atomic charges were obtained. In addition, PM3-Mulliken charge analyses were carried out with *Gaussian03* and MOPAC 5.011mn as part of a comparison between partial atomic charges assigned to optimized geometries calculated by PM3 and PDDG/PM3.

The MM force fields employed are AMBER,⁵⁵ AMBER*, CVFF,⁵³ CFF91,⁸⁹ MMFF94,⁵⁶ OPLS2005,⁵⁷ and Tripos.⁵⁴ The AMBER force field employed is the ff03 version,⁹⁰ in conjunction with the general atom force field⁹¹ (GAFF) commonly utilized for small organic systems. The AMBER* force field contains additional atomic parameters as implemented in *MacroModel*. In most cases, we elected to use nonrigid water with each force field's default parameters for water molecules. However, the AMBER* force field was locally modified to use OPLS2005 nonrigid water (in OPLS2005, the nonrigid water has the same Lennard-Jones parameters as the rigid TIP3P water model), and the AMBER force field, via the SOLVATEOCT command, utilized the rigid TIP3P water model. Stretch, bend, Coulombic, and Lennard-Jones parameters for the water models used can be found in each of the force field's descriptions (see references above), except for AMBER*, for which the modified water parameters are described by the OPLS2005 reference.

In addition to the standard force fields, we also employ local modifications of the force fields that substitute CM4 charges for their default partial charges. This combination of a force field and CM4 charges is denoted as X-CM4, where X is the name of the original force field. In contrast to gas-phase SE-MO calculations, when CMx charges are used with MM, they can and do alter the optimized geometry. Note that when we use CM4 charges with MM calculations, we use gas-phase M05-2X/6–31+G(d,p)/CM4 charges calculated at gas-phase M05-2X/6–31+G(d,p) geometries for gas-phase MM calculations, and we use SM6/M05-2X/6–31+G(d,p)/CM4 aqueous-phase charges calculated at aqueous-phase SM6/M05-2X/6–31+G(d,p) geometries for aqueous-phase MM calculations.

General MM optimization conditions consisted of at least 1000 steps of conjugate gradient minimization with an energy gradient cutoff of $0.01 \text{ kcal mol}^{-1} \text{ Å}^{-1}$. In cases where water was included, the entire system was minimized, and although the stringent conditions for termination we have established were not typically reached, the system energy fluctuated only slightly around a stable potential energy. To test the effect of water position around the small molecules on geometries, six random orientations of 2-AMP and 2,4-DAP within the water box were used in minimizations with OPLS2005-CM4 and MMFF94-CM4. Because this process yielded only a slight standard deviation in geometries (on the order of less than 0.006 Å in bond length and 0.7 degrees in bond angle), we used only the default water orientation for all other aqueous-phase MM optimizations. We note, however, the

2-AMP and 2,4-DAP bond lengths and angles used for OPLS2005-CM4 and MMFF94-CM4 assessment represent the average of these six results. In binding energy studies, water molecules were placed at positions on 2-AMP, 1H-2-AMP, 2,4-DAP, and 1H-2,4-DAP, where the DHFR-MTX crystal structure denoted that hydrogen bonding takes place between the ligand and the enzyme.⁴⁵ The small molecule and water were optimized together, and the binding energy of the complex computed by subtracting the energies of the individual optimized molecules.

While testing the MM force fields, it was found that the CVFF and CFF91 force fields do not properly assign partial atomic charges to the 1H-2-AMP or 1H-2,4-DAP cations because the total charge assigned to the molecule is not +1.0. We therefore calculated the gas-phase Hartree–Fock molecular electrostatic potential with the 6–31+G(d,p) basis set for the neutral and cationic species and obtained ChelpG⁸⁰ electrostatic-potential-filling charges to substitute into the force fields. Geometry optimizations with the ChelpG partial charges were then carried out in both the gaseous and aqueous phases, and the resulting geometries, denoted CVFF-HF and CFF91-HF, were used in our evaluation.

2.2. Platforms, Software, and Molecules. Quantum mechanical calculations (WFT, DFT, and SE-MO) were performed on an IBM Power4 (p690 and p655) computer system running under the AIX operating system and an SGI Altix cluster running under the Linux operating system. Molecular mechanics calculations were performed on a Silicon Graphics O2 workstation running under the IRIX 6.5 operating system. Molecules were constructed for quantal calculations using the *GaussView 3.0* (Gaussian, Inc.) visualization program, and the generated Z-matrices were converted to Cartesian coordinates where appropriate. Molecules for MM calculations were constructed using *InsightII 2005* (Accelrys, Inc.) for the CFF91 and CVFF force fields, SYBYL 7.3 (Tripos, Inc.) for the MMFF94, AMBER and Tripos force fields, and *Maestro 7.5* (Schrodinger, Inc.) for the MMFF94, AMBER*, and OPLS2005 force fields (utilizing the *MacroModel* and *Impact* applications, respectively). These programs were also used to set up and run the MM minimizations except for AMBER minimizations, for which SYBYL was used only to generate molecular coordinates in Protein Data Bank or Mol2 formats, and the AMBER 9⁹² suite was used for the minimizations. The small molecules included in the present study are illustrated in Figure 3. Each molecule was modeled in both its neutral and protonated form. When an exocyclic amine is present, the proton was added there. Otherwise, the proton was added on a heterocyclic amine.

3. Results and Discussion

3.1. Error Analysis. To rank the methods we chose to test, we calculated the unsigned residual between a calculated value with method *m* and the corresponding benchmark value

$$R_i^{x,y,m} = |x_i^{\text{calcd},m}(y) - x_i^{\text{benchmark}}(y)| \quad (1)$$

where *y* is the phase (gas phase or aqueous phase) and *x_i(y)* signifies case *i* of molecular property *x* in phase *y*, for example, when *x* is bond length (*r*), *x₁(y)* is the first bond

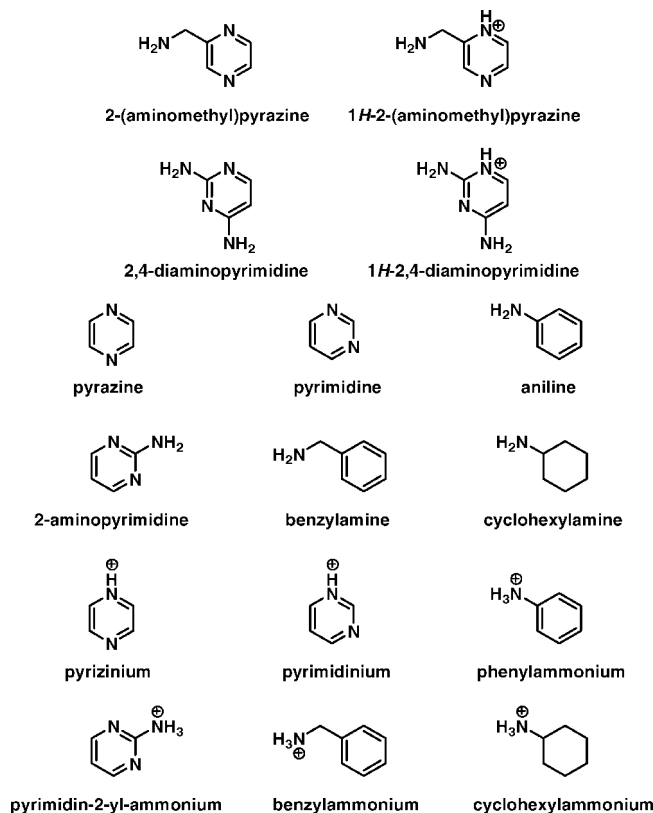


Figure 3. Structures of small molecules used in current study.

length. Alternatively, *x* could stand for partial charge (*q*), bond angle (*θ*), or binding energy (*E_b*). The overall error in a particular molecular property *x* for a particular method *m* and phase *y* is quantified by the mean unsigned error (MUE)

$$\text{MUE}_{x,y,m} = \frac{\sum_{i=1}^{n_{x,y}} R_i^{x,y,m}}{n_{x,y}} \quad (2)$$

where *n_{x,y}* is the number of combinations of *x_i* and *y* for which *R_i^{x,y}* is evaluated.

We also calculate the average mean unsigned error (AMUE) for each property across *N* methods

$$\text{AMUE} = \frac{\sum_{m=1}^N \text{MUE}_{x,y,m}}{N} \quad (3)$$

Division of this value by a method's MUE yields the reduced (unitless) mean unsigned error (RMUE) for the method for partial charge, bond length, or bond angle.

$$\text{RMUE}_{x,y,m} = \frac{\text{MUE}_{x,y,m}}{\text{AMUE}_{x,y}} \quad (4)$$

The RMUE is a measure of each method's performance relative to the mean of the others for calculating a particular molecular property. A value of 1.0 indicates that the method is average. Lower values indicate better methods, while higher values indicate worse methods. The reason for introduction of these unitless reduced quantities is so that we can combine errors for *r* and *θ* (which have different

Table 1. Mean Unsigned Error (Å and deg) for Gas-Phase Bond Lengths and Angles Between Those Calculated with CCSD and Each Functional

method	2-AMP		1H-2-AMP		2,4-DAP		1H-2,4-DAP	
	bond length	bond angle	bond length	bond angle	bond length	bond angle	bond length	bond angle
B3LYP	0.003	0.73	0.003	0.37	0.003	0.82	0.003	0.27
MPW1PW	0.005	0.44	0.005	0.49	0.005	0.87	0.004	0.26
mPWB1K	0.011	0.52	0.010	0.42	0.012	0.99	0.010	0.26
mPW1KCIS	0.005	0.78	0.005	0.50	0.004	0.90	0.004	0.28
M06-L	0.006	0.52	0.005	0.33	0.004	0.54	0.004	0.22
M05-2X	0.004	0.28	0.004	0.26	0.005	0.94	0.003	0.22
AMUE ^a	0.006	0.54	0.005	0.40	0.006	0.84	0.005	0.25

^a Mean of entire column.

units) to make an overall assessment for combined geometric performance.

We define a reduced deviance ($D_{y,m}$) of a SE-MO or MM method from the average performance in either the gas or aqueous phase by averaging the RMUE for r and θ in each method for both 2-AMP and 1H-2-AMP

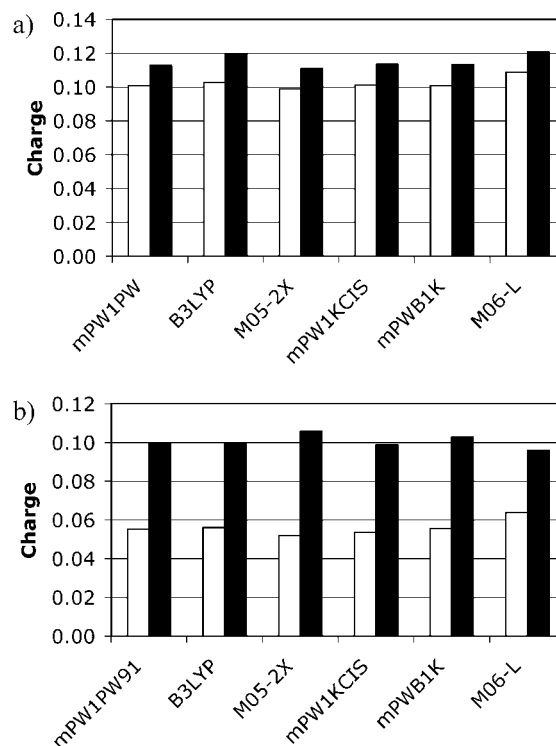
$$D_{y,m} = \frac{1}{4}(\text{RMUE}_{r,y,m}^{2\text{-AMP}} + \text{RMUE}_{\theta,y,m}^{2\text{-AMP}} + \text{RMUE}_{r,y,m}^{1\text{H-2-AMP}} + \text{RMUE}_{\theta,y,m}^{1\text{H-2-AMP}}) \quad (5)$$

The reduced deviance for 2,4-DAP and 1H-2,4-DAP is averaged with the reduced deviance for 2-AMP and 1H-2-AMP to yield an overall performance. Reduced deviance in partial charge (q) assignment is calculated similarly. Reduced deviance for the validation of the most accurate methods is also calculated similarly, with all molecules taken into account.

3.2. Establishing Geometry and Partial Charge Benchmark Sets. The MUE of the gas-phase geometries calculated by DFT with respect to the CCSD-calculated geometries is given in Table 1. All DFT methods perform well, with mean unsigned errors within ~ 0.01 Å for bond length and less than one degree for bond angles. On the basis of its high degree of accuracy for both bond length and angle, we chose to proceed with the M05-2X level of theory to generate our geometric benchmark set.

We selected the well-validated^{51,59,93–96} CM4 charge model to obtain benchmark partial atomic charges. To examine whether the CM4 charge model would assign partial charge similarly for each functional used, we compared the gas-phase CCSD partial charges generated by Mulliken population analysis (a class II⁴⁶ charge method) to CM4 partial charges assigned by each density functional tested (note that the CM4 charges are probably more accurate). The results are summarized in Figure 4. In this figure and in this whole article, charges are given in atomic units, in which the charge on a bare proton is 1.0. Overall, the mean unsigned deviations between the CCSD Mulliken analysis and the CM4-assigned partial charges vary by ≤ 0.01 charge units regardless of the functional or phase tested. We therefore applied the geometrically accurate M05-2X functional in conjunction with M05-2X/CM4 partial charges to obtain our benchmark set.

3.3. Exploration of CVFF and CFF91 Atom Typing and Charge Distribution. As mentioned in section 2.2, a deficiency in the CVFF and CFF91 force fields is that they

**Figure 4.** Mean unsigned deviation of DFT/CM4 charges relative to CCSD/Mulliken charges for (a) 2-AMP and its cation (white, 2-AMP; black, 1H-2-AMP) and (b) DAP and its cation (white, DAP; black, 1H-DAP).

do not assign a total charge of +1.0 to the pyrazinium or pyridinium cations. Upon further exploration using 2-AMP, the default atom type assigned to N1 in 2-AMP by both force fields is “np” (an sp^2 nitrogen in a 5- or 6-membered ring), and the partial charge on this unprotonated nitrogen is -0.22 in CVFF and -0.48 in CFF91. Upon protonation and automated reassignment of atom types by *InsightII 2005*, the CVFF nitrogen atom type remains unchanged, and the CFF91 atom type changes to “nh+” (a protonated nitrogen in a 6-membered ring). The protons added have partial charges of $+0.28$ (CVFF) and $+0.33$ (CFF91), and the partial charges on the nitrogens change to -0.50 and -0.81 charge units, respectively. The partial charge on all other atoms in the molecule are unaffected by the addition of the proton. This charge balancing yields a total charge on the molecule of 0, which is incorrect for a cation.

Some of the partial atomic charges in the CVFF and CFF91 force fields are derived from fits to the Hartree–Fock molecular electrostatic potential,^{97–99} and we took this as a cue for how to correct the problem in a way consistent with these force fields. In particular, we carried out single-point, gas-phase Hartree–Fock (6–31G(d,p) basis set) calculations on the optimized 2-AMP and 1H-2-AMP geometries and obtained partial atomic charges by electrostatic potential fitting with the ChelpG⁸⁰ algorithm. The resulting partial atomic charges were used in the CVFF and CFF91 force fields for geometry optimizations in both the gas and aqueous phases. We used the gas-phase partial charges in both phases because the original CVFF and CFF91 partial charges are based on gas-phase calculations (we note that all MM force fields considered in this article use partial charges that do

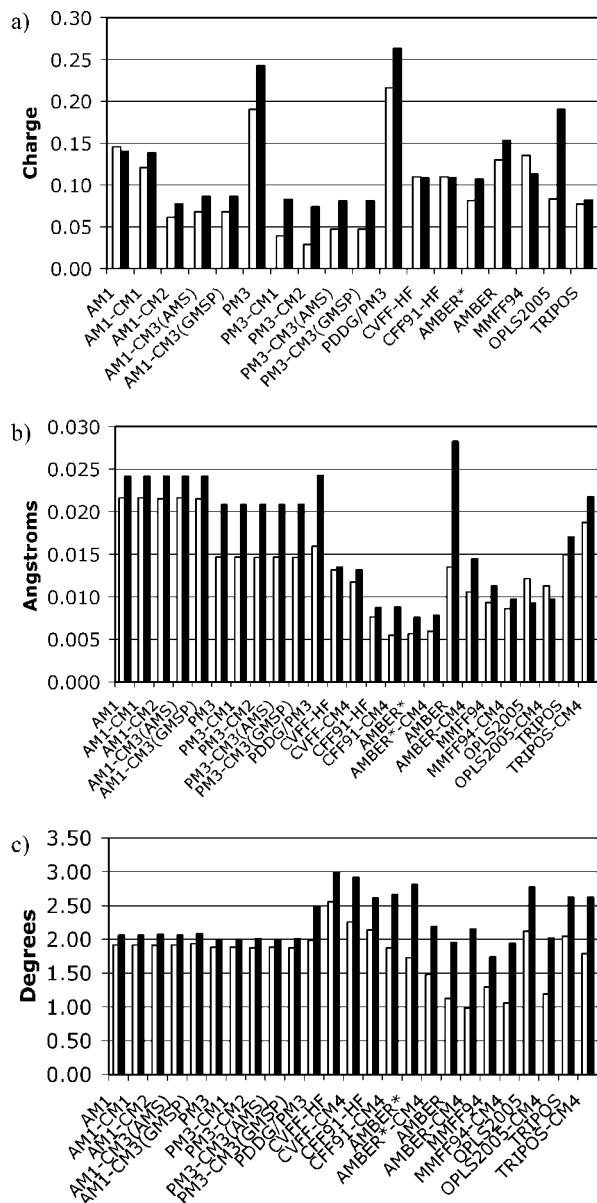


Figure 5. MUE in (a) partial charge, (b) bond length, and (c) bond angle for selected SE-MO and MM methods in the gas phase relative to M05-2X/CM4 (white, 2-AMP; black, 1H-2-AMP). AMS denotes AMSOL, and GMSP denotes GAMESS-PLUS.

not depend on the phase). To denote the new Hartree–Fock partial charges in the force fields, we name the force fields that use these newly assigned charges as CVFF-HF and CFF91-HF.

3.4. Evaluation of SE-MO and MM Calculations. We tested available SE-MO and MM parameter sets in an effort to select the most accurate method for modeling our small molecules both in the gas phase and in solvent. In addition, we substituted CM4 charges into the MM force fields tested to observe effects on geometric accuracy. A summary of the results of these tests is given in Figures 5 and 6 (gas phase) and 7 and 8 (aqueous phase). In the gas phase, the AM1 and PM3 SE-MO methods utilizing the CM1, CM2, and CM3 charge models, as well as the CVFF-HF, CFF91-HF, AMBER*, and MMFF94 force fields all predict the partial charges of the atoms in both sets of molecules to within 0.15

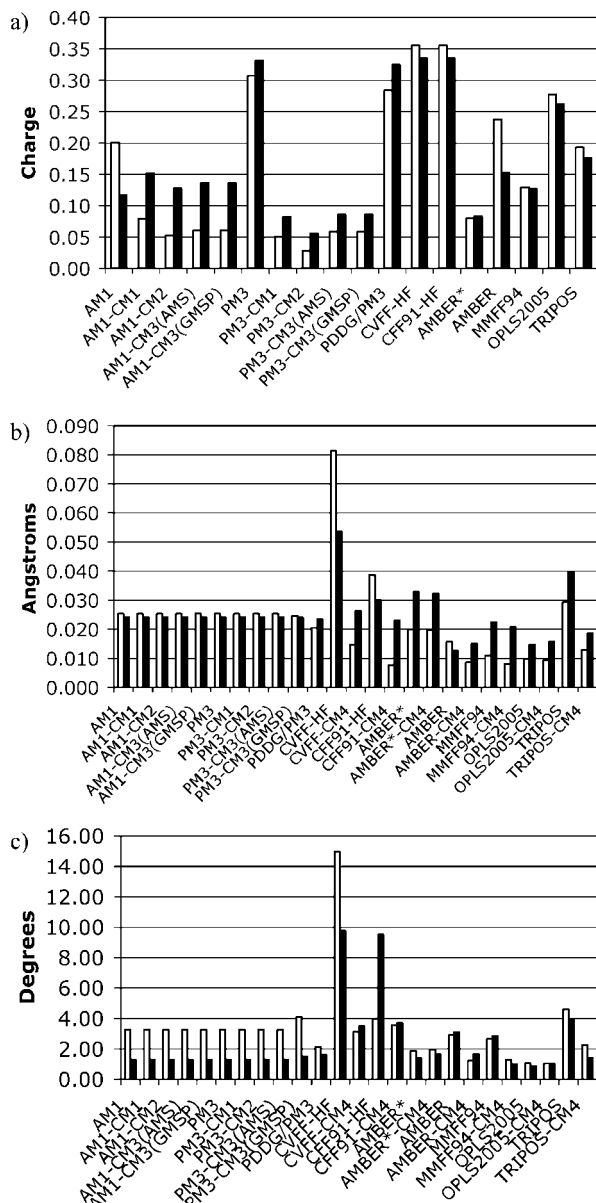


Figure 6. MUE in (a) partial charge, (b) bond length, and (c) bond angle for selected SE-MO and MM methods in the gas phase relative to M05-2X/CM4 (white, 2,4-DAP; black, 1H-2,4-DAP).

(Figures 5a and 6a). In most of the methods tested, the partial charge assignment becomes less accurate in the cationic species: the most notable examples, for which the average error increases by a factor of 2, are the OPLS2005 force field for 1H-2-AMP, the PM3-CM1, -CM2, and -CM3 methods for 1H-2-AMP, and the SE-MO methods utilizing the CM1, CM2, and CM3 charge models for 1H-2,4-DAP (with the exception of PM3-CM3). In contrast, the MMFF94 force field becomes more accurate for 1H-2-AMP, and the AMBER force field becomes more almost 2-fold more accurate for 1H-2,4-DAP, although AMBER's overall charge assignment is somewhat inaccurate (the mean MUE for the neutral and cationic species together is 0.20). The force fields with CM4 charges substituted are not included in the partial charge analysis since their mean unsigned error in partial charge is always zero.

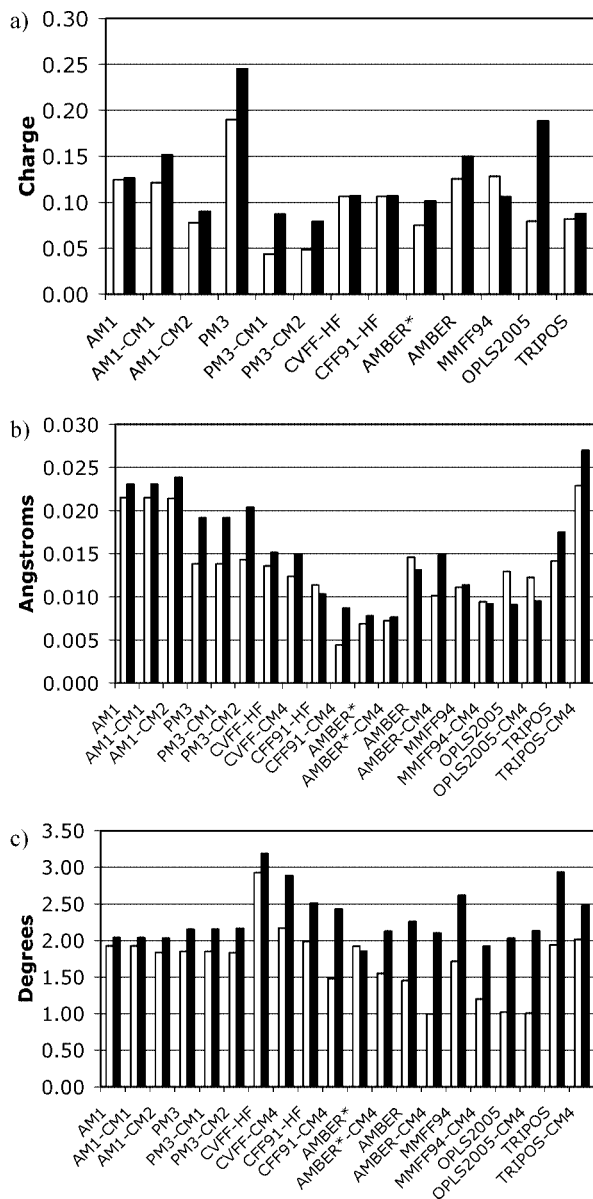


Figure 7. MUE in (a) partial charge, (b) bond length, and (c) bond angle for selected SE-MO and MM methods in solution relative to M05-2X/CM4 (white, 2-AMP; black, 1H-2-AMP).

The gas-phase partial charges assigned by the PDDG/PM3 method have a mean unsigned error of 0.22 and 0.28 for the neutral species and 0.26 and 0.32 for the cationic species of 2-AMP and 2,4-DAP, respectively. Table 2 summarizes our comparison of the PDDG/PM3 charges to Mulliken population analysis of the PM3-optimized structure of 2-AMP as calculated by both *Gaussian03* and MOPAC 5.011mn (the numbering system is given in Figure S1 of the Supporting Information); this comparison verifies the consistency of the two programs (as a check). Figures 5 and 6 show that PDDG/PM3 is less accurate than PM3 for gas-phase partial charges, bond lengths, and bond angles for both 2-AMP and 1H-2-AMP and for gas-phase bond angles for both 2,4-DAP and 1H-2,4-DAP. Because the PDDG reparameterization of PM3 deteriorated the performance for these molecules, PDDG/PM3 was not considered in the aqueous-phase calculations that follow.

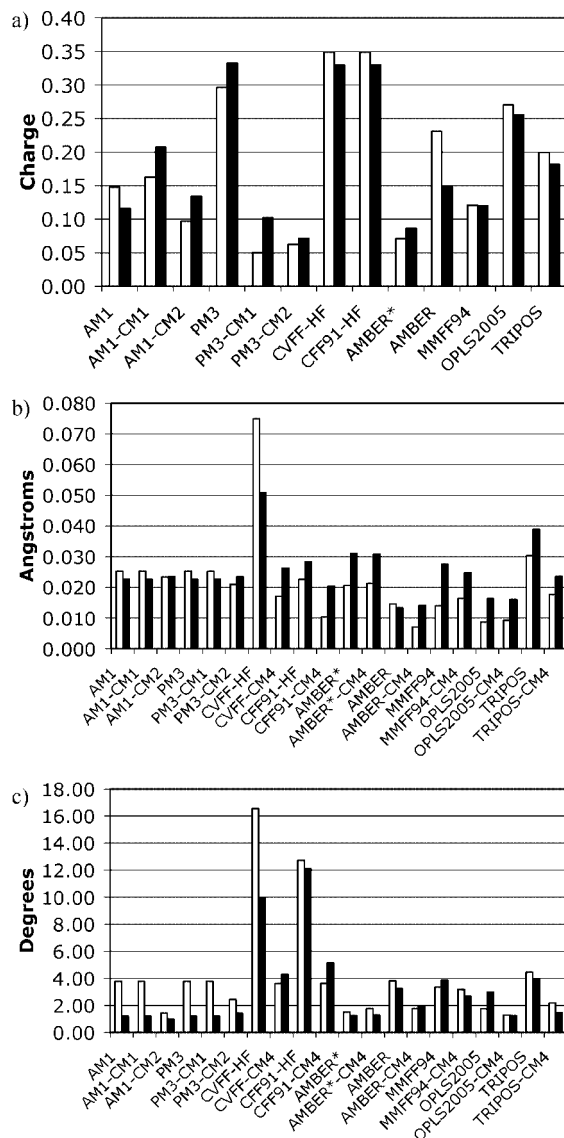


Figure 8. MUE in (a) partial charge, (b) bond length, and (c) bond angle for selected SE-MO and MM methods in solution relative to M05-2X/CM4 (white, 2,4-DAP; black, 1H-2,4-DAP).

Table 2. Gas-Phase Partial Charges of 2-AMP Calculated by PM3 (Mulliken Population Analysis) and PDDG/PM3

atom	PM3(G03)	PM3(MOPAC)	PDDG/PM3(MOPAC)
C1	-0.1310	-0.1309	-0.1316
C2	-0.1067	-0.1067	-0.1647
N3	-0.0382	-0.0382	-0.0274
C4	-0.1095	-0.1094	-0.1664
C5	-0.1085	-0.1085	-0.1581
N6	-0.0239	-0.0239	-0.0211
H7	0.1344	0.1344	0.1814
H8	0.1319	0.1319	0.1798
H9	0.1316	0.1316	0.1789
C10	-0.0569	-0.0569	-0.1491
H11	0.0677	0.0677	0.1151
H12	0.0830	0.0830	0.1296
N13	-0.0282	-0.0282	-0.0928
H14	0.0301	0.0301	0.0665
H15	0.0242	0.0242	0.0598

For 2-AMP and 1H-2-AMP, gas-phase bond lengths (Figure 5b) are predicted to within ~ 0.01 Å only by the

CFF91-HF, AMBER*, MMFF94, and OPLS2005 force fields. All methods except OPLS2005 become less accurate when modeling the cationic species. Notably, the AMBER force field becomes 2-fold less accurate when modeling the cation, and it clearly benefits from revised charges because the AMBER-CM4 force field yields a 2-fold performance improvement when modeling bond lengths. On average, force fields with CM4 charges substituted tend to perform slightly better than those with default charges, except in the case of TRIPOS-CM4. Gas-phase bond angles (Figure 5c) are predicted to within ~ 2.0 degrees by all the SE-MO methods, while the majority of the MM force fields are ~ 0.5 – 1° less accurate. Exceptions include the MMFF94 and AMBER force fields, which predict them to within $\sim 1.5^\circ$. All methods are slightly more inaccurate dealing with the cation. As with bond lengths, when CM4 charges are substituted, bond angle predictivity is increased slightly in most methods. OPLS2005-CM4 is notable in that it has a 2-fold improvement in performance over OPLS2005 when modeling the neutral species and a smaller, yet significant ($\sim 0.5^\circ$), increase in accuracy for the cationic species.

For 2,4-DAP and 1*H*-2,4-DAP, bond lengths (Figure 6b) are predicted to an accuracy of ~ 0.025 Å by all methods except CVFF-HF, CFF91-HF, and TRIPOS. The MMFF94 and OPLS2005 force fields both display accuracy less than 0.01 Å for 2,4-DAP, while AMBER and OPLS2005 are accurate to better than 0.015 Å for 1*H*-2,4-DAP. MMFF94 is 2-fold less accurate for the 1*H*-2,4-DAP cation. With regard to 2,4-DAP and 1*H*-2,4-DAP gas-phase bond angles, the best performing methods include the AMBER*, AMBER, MMFF94, and OPLS2005 force fields, predicting bond angle to within 1.9° , 3.1° , 2.9° , and 1.1° for both 2,4-DAP and 1*H*-2,4-DAP, respectively. With respect to force fields with CM4 charges substituted, the clearest example of a force field benefiting from revised charges is the CVFF-HF and CFF91-HF force fields which, as previously discussed, have major partial charge assignment problems. When CM4 charges are substituted into these force fields, both methods show a dramatic improvement in performance for modeling bond lengths, with CVFF-CM4 yielding a 4-fold more accurate bond length prediction overall, and CFF91-CM4 yielding the same improvement in accuracy when modeling the neutral species. The TRIPOS force field also benefits, with TRIPOS-CM4 displaying a 2-fold improvement in accuracy for both molecules.

All methods except CVFF-HF, CFF91-HF, and TRIPOS predict bond angles to within ~ 4.0 degrees for both species (Figure 6c). Both CVFF-HF and CFF91-HF presented a challenge when attempting to minimize structure with respect to the fact that 1*H*-2,4-DAP could not be assigned proper atomic partial charges. As described earlier, we substituted the default charges with ChelpG charges derived from HF calculations on 2,4-DAP and 1*H*-2,4-DAP. When these charges were substituted into the force field, the charges on the protons on the exocyclic amines were made largely more positive than the default charges. This effect, in conjunction with the largely negatively charged heterocyclic amines, caused the minimization to distort the sp^3 structure of the exocyclic amines and pull the protons toward the pyrimidine

ring. The result is a large error in the assignment of bond angle. When the minimization is performed with the default charges, no distortion of bond angle takes place; however, these charges correspond to a nonphysical total molecular charge. As observed in tests already described, CM4 charge substitution greatly improves geometric modeling performance. The CVFF-CM4 and CFF91-CM4 force fields essentially eliminate the problems seen with these force fields, causing both to model bond angles with accuracy on par with the rest of the methods used. In addition, the AMBER-CM4, MMFF94-CM4, and TRIPOS-CM4 force fields all yield accuracies at least 2-fold greater than their counterparts.

For 2-AMP and 1*H*-2-AMP in the aqueous phase, the AM1-CM2, PM3-CM1, and PM3-CM2 methods, as well as the CVFF-HF, CFF91-HF, AMBER*, MMFF94, and TRIPOS force fields retain a high degree of accuracy for partial charge prediction ($MUE \leq 0.13$, Figure 7a). Among these methods, the MMFF94 force field is again the only method to become more accurate when modeling the protonated species. On the other hand, the OPLS2005 force field becomes 2-fold less accurate when predicting the charge of the cation. Bond lengths (Figure 7b) are calculated with an error similar to that of the gas phase, with the CFF91-HF, AMBER*, MMFF94, and OPLS2005 force fields maintaining a mean unsigned error of ~ 0.01 Å. Interestingly, while it becomes much less accurate when assigning partial charge to the ionic species, the OPLS2005 force field becomes more accurate in bond length prediction. As with 2-AMP and 1*H*-2-AMP in the gas phase, CM4 partial charge substitution generally produces a modest increase in bond length accuracy. A notable exception is the CFF91-CM4 force field, with an $MUE < 0.005$ Å. When compared to calculations in the gas phase, all methods either retain their accuracy or become slightly less accurate when predicting bond angle in solution except for the OPLS2005 force field. Except for the CVFF-HF, AMBER*, and TRIPOS force fields, all methods are still accurate to $\sim 2.5^\circ$ or less (Figure 7c). Again, on average, CM4 charge substitution slightly increases bond angle accuracy for both species.

In the aqueous-phase treatment of 2,4-DAP and 1*H*-2,4-DAP, partial charge performance is quite varied (Figure 8a). The best performing methods are the PM3-CM1 and -CM2 SE-MO methods and the AMBER* and MMFF94 force fields, which all model the partial charges to an accuracy of 0.12 or better. In the treatment of these molecules, all the force fields except AMBER* become more accurate to varying degrees when dealing with the cationic species. The quality of modeling the bond lengths (Figure 8b) is similar for most methods; however, the CVFF-HF force field is particularly inaccurate, with a MUE in bond length for 2,4-DAP of 0.075 Å. The AMBER and OPLS2005 force fields perform particularly well for both species, with MUE s of ≤ 0.016 Å overall. However, the OPLS2005 force field becomes 2-fold less accurate when modeling the cationic species. MMFF94 also performs relatively well, modeling the neutral species with a $MUE \leq 0.014$ Å, but, like OPLS2005, becomes 2-fold less accurate when modeling the cation. CVFF-CM4, CFF91-CM4, AMBER-CM4 (for the

Table 3. Reduced Deviance ($D_{y,m}$) in Partial Charge for Each SE-MO and MM Method Tested

method	2-AMP and 1H-2-AMP			2,4-DAP and 1H-2,4-DAP			all molecules		
	gas	aqueous	mean	gas	aqueous	mean	gas	aqueous	mean
PM3-CM2	0.45	0.56	0.50	0.25	0.36	0.31	0.35	0.46	0.40
PM3-CM1	0.54	0.56	0.55	0.40	0.41	0.40	0.47	0.49	0.48
AM1-CM2	0.63	0.75	0.69	0.54	0.62	0.58	0.58	0.68	0.63
AMBER*	0.85	0.78	0.81	0.49	0.42	0.46	0.67	0.60	0.64
MMFF94	1.15	1.06	1.11	0.77	0.65	0.71	0.96	0.85	0.91
TRIPOS	0.73	0.75	0.74	1.12	1.03	1.07	0.92	0.89	0.91
AM1-CM1	1.18	1.21	1.19	0.69	1.00	0.84	0.93	1.10	1.02
AM1	1.32	1.12	1.22	0.97	0.71	0.84	1.14	0.92	1.03
AMBER	1.29	1.22	1.25	1.19	1.02	1.11	1.24	1.12	1.18
OPLS2005	1.20	1.15	1.17	1.63	1.42	1.52	1.41	1.28	1.35
CVFF-HF	1.00	0.96	0.98	2.09	1.83	1.96	1.54	1.39	1.47
CFF91-HF	1.00	0.96	0.98	2.09	1.83	1.96	1.54	1.39	1.47
PM3	1.96	1.92	1.94	1.92	1.70	1.81	1.94	1.81	1.88
PM3-CM3(GMSP) ^{b,a}	0.57	ND	ND	0.43	ND	ND	0.50	ND	ND
PM3-CM3(AMS) ^{b,a}	0.57	ND	ND	0.43	ND	ND	0.50	ND	ND
AM1-CM3(GMSP) ^{b,a}	0.70	ND	ND	0.58	ND	ND	0.64	ND	ND
AM1-CM3(AMS) ^{b,a}	0.70	ND	ND	0.58	ND	ND	0.64	ND	ND
PDDG/PM3 ^a	2.17	ND	ND	1.83	ND	ND	2.00	ND	ND

^a ND = not determined. ^b AMS = calculated using AMSOL7.1, GMSP = calculated using GAMESPLUSS.

Table 4. Reduced Deviance ($D_{y,m}$) in Combined Geometry for Each SE-MO and MM Method Tested

method	2-AMP and 1H-2-AMP			2,4-DAP and 1H-2,4-DAP			all molecules		
	gas	aqueous	mean	gas	aqueous	mean	gas	aqueous	mean
OPLS2005-CM4	0.73	0.76	0.75	0.45	0.45	0.45	0.59	0.60	0.60
AMBER-CM4	0.78	0.80	0.79	0.52	0.49	0.50	0.65	0.65	0.65
MMFF94-CM4	0.66	0.71	0.68	0.50	0.85	0.68	0.58	0.78	0.68
OPLS2005	0.96	0.76	0.86	0.43	0.61	0.52	0.69	0.68	0.69
AMBER*-CM4	0.67	0.71	0.69	0.86	0.77	0.82	0.77	0.74	0.75
AMBER*	0.76	0.73	0.75	0.84	0.74	0.79	0.80	0.74	0.77
CFF91-CM4	0.78	0.70	0.74	0.97	0.96	0.97	0.88	0.83	0.85
MMFF94	0.70	0.93	0.81	0.85	0.95	0.90	0.78	0.94	0.86
AMBER	1.03	0.94	0.98	0.85	0.80	0.82	0.94	0.87	0.90
PM3-CM2	1.05	1.10	1.08	0.91	0.74	0.83	0.98	0.92	0.95
PM3	1.05	1.07	1.06	0.91	0.85	0.88	0.98	0.96	0.97
PM3-CM1	1.05	1.07	1.06	0.91	0.85	0.88	0.98	0.96	0.97
TRIPOS-CM4	1.19	1.43	1.31	0.65	0.69	0.67	0.92	1.06	0.99
AM1-CM2	1.24	1.27	1.25	0.91	0.68	0.79	1.07	0.98	1.02
CVFF-CM4	1.04	1.10	1.07	1.03	1.03	1.03	1.04	1.06	1.05
AM1-CM1	1.24	1.28	1.26	0.91	0.85	0.88	1.07	1.06	1.07
AM1	1.24	1.28	1.26	0.91	0.85	0.88	1.07	1.06	1.07
TRIPOS	1.09	1.15	1.12	1.49	1.33	1.41	1.29	1.24	1.26
CFF91-HF	0.85	0.94	0.90	2.02	2.30	2.16	1.44	1.62	1.53
CVFF-HF	1.12	1.27	1.20	3.60	3.20	3.40	2.36	2.24	2.30
PM3-CM3(AMS) ^{b,a}	1.05	ND	ND	0.91	ND	ND	0.98	ND	ND
PDDG/PM3 ^a	1.19	ND	ND	0.79	ND	ND	0.99	ND	ND
PM3-CM3(GMSP) ^{b,a}	1.05	ND	ND	0.99	ND	ND	1.02	ND	ND
AM1-CM3(GMSP) ^{b,a}	1.24	ND	ND	0.91	ND	ND	1.07	ND	ND
AM1-CM3(AMS) ^{b,a}	1.24	ND	ND	0.91	ND	ND	1.07	ND	ND

^a ND = not determined. ^b AMS = calculated using AMSOL7.1, GMSP = calculated using GAMESPLUSS.

neutral species) and TRIPOS-CM4 all yield about a 2-fold increase in accuracy. Other force fields with CM4 charges in place show little to no difference. Bond angles (Figure 8c) are treated with an accuracy similar to that in the gas phase, with the CVFF-HF and CFF91-HF force fields performing very poorly ($\text{MUE} \geq 10$ degrees). All methods except the TRIPOS, CVFF-CM4, and CFF91-CM4 force fields model bond angles to within $\sim 4.0^\circ$, although CVFF-CM4 and CFF91-CM4 again represent a great improvement over their parent force fields. Again, CM4 treatment of the force fields generally leads to an almost 2-fold increase in accuracy for at least one, if not both, of the molecules modeled by each method.

Comparison of Figures 5 and 6 to Figures 7 and 8 shows that the SE-MO methods and force fields are about equally accurate when comparing gas-phase to aqueous-phase results, so no one method in particular stands out as extremely ill-suited to work in either the gaseous or aqueous phase.

3.5. Overall Geometric Assessment. To make an overall geometric assessment, we consider the reduced deviances ($D_{y,m}$) in partial charge assignment (Table 3) and geometric modeling (Table 4). Reduced deviance in partial charge shows that seven of the eleven SE-MO methods tested (the exceptions being AM1, AM1-CM1, PM3, and PDDG/PM3) perform better than the average method. In fact, PM3-CM2 is a factor of 2.5 better than average. The only MM methods

that predict partial charge better than average are the AMBER*, MMFF94, and TRIPOS force fields. This assessment, however, also must take into account that the PM3-CMx and AM1-CMx methods use charge methods that serve as precursors to charge model 4. Since the training sets for the various CMx models share some molecules, it is perhaps not surprising that the partial atomic charges of the various CMx models show some agreement.

Across the molecules, reduced deviance is fairly consistent when the large errors in CVFF-HF and CFF91-HF for 2,4-DAP and 1*H*-2,4-DAP are taken into account. A slight skewing of the data may be occurring because of the aforementioned error because CVFF-HF and CFF91-HF actually perform as well as the average in partial charge assignment for 2-AMP and 1*H*-2-AMP but perform so poorly in the other cases that their overall reduced deviance is high. However, if we were to carry on with either the CVFF-HF or CFF91-HF force fields by using the original method⁹⁸ of partial charge assignment for these force fields, we would be forced to perform electrostatic potential fitting on each molecule we study, and it is known that one of the weaknesses of ChelpG is fitting to the molecular electrostatic potential (MEP) of larger systems, where buried atoms are screened from the points where the MEP is evaluated, and changes of the partial charges on these atoms have only small effects on the MEP. Attempting to fit these charges, then, often produces nonphysical results.^{80–83}

Whereas the SE-MO charges and CMx charges are different in the gas phase and the aqueous phase, the MM charges (with the exception of those models using CM4 charges) are the same. Table 3 allows for further examination of an issue discussed briefly at the end of section 3.4, namely, the question of whether standard MM charges are more appropriate for the gas phase or for liquid-phase solution. Some force fields are explicit on this issue, for example, OPLS is explicitly named for its use in liquid-phase simulations. Others are implicitly designed for use in liquid phases simply because that is where the greatest number of applications of MM force fields occur. Table 3 shows that MMFF94, AMBER, OPLS2005, CVFF-HF, and CFF91-HF all perform significantly better for partial atomic charges in water than in the gas phase, and AMBER* and TRIPOS are slightly better in the aqueous phase than in the gas phase. The finding that the partial charges are more indicative of the charge distribution in the aqueous phase than the gas phase in all seven cases is quite remarkable and is encouraging.

The reduced deviance in geometric modeling is given in Table 4. Of particular interest are the excellent performance of the MM-CM4 methods and the repair of the CFF91-HF force field. Prior to partial charge replacement, the CFF91-HF force field has a geometric $D_{y,m}$ equal to 1.54. After CM4 treatment, the $D_{y,m}$ for CVFF-CM4 is 0.86, representing a significant increase in the geometric modeling capabilities of the force field by fixing partial charge assignment problems. Other MM-CM4 methods also perform very well, with OPLS2005-CM4, AMBER-CM4, and MMFF94-CM performing 39%, 36%, and 32% better than the average, respectively. As far as non-CM4 treated force fields are concerned, the OPLS2005 force

field performs the best ($D_{y,m} = 0.68$), with AMBER*, MMFF94, and AMBER also modeling geometries better than average. The tested SE-MO methods all perform very similarly around a $D_{y,m}$ of ~ 1.0 .

Another interesting point in Table 4 is the comparison of performance in the gas phase with that in aqueous solution. Of the fourteen MM rows in Table 4, six show smaller errors for gas-phase geometries, and eight show smaller errors for aqueous geometries (of the eight other methods for which such comparison is possible, all show better agreement in the aqueous phase). The MMFF94 method is particularly noteworthy in showing (with either MMFF94 charges or CM4 charges) much better accuracy for gas-phase geometries than for aqueous ones.

On the basis of a combined assessment of partial charge assignment and geometric modeling performance, we find that the AMBER, AMBER*, MMFF94, and OPLS2005 force fields, along with their CM4-treated counterparts, would be suitable for carrying on into a validation step against a larger set of molecules. Each force field has some particular advantages. The AMBER force field has traditionally been highly regarded for its use in modeling biopolymers, and in this case, is used with the incorporated parameters of the General Amber Force Field (GAFF). Atom types in GAFF are designed to be more general than those of traditional AMBER force fields, in an effort to cover a larger portion of organic space. The parametrization of GAFF was developed in an effort to reproduce restrained electrostatic potential (RESP) charges^{81,85} at the MP2/6–31G(d) level and reproduce MP2/6–31G(d) and crystallographic geometries.⁹¹ The AMBER* force field has been implemented in *MacroModel* and been modified to better reproduce HF/6–31+G(d) data on peptides, as well as small organic molecules, especially those with nitrogen as a component.^{100,101} The MMFF94 force field has been parametrized against a large (~ 2800 structures), high-quality ab initio training set for use with both organic molecules and biopolymers, specifically in solution.¹⁰²

3.6. Comparison of Binding Energies. We carried out gas-phase binding energy calculations between 2-AMP, 1*H*-2-AMP, 2,4-DAP, and 1*H*-2,4-DAP and water molecules placed at locations where hydrogen bonding has been described in the DHFR-MTX crystal structure (Figure 9). The results of this study are presented in Figure 10, wherein the MUE is the mean unsigned error for a method in binding energy for all neutral molecules, all charged molecules, or both sets together. Compared to the M05-2X benchmark binding energies, all methods except for CVFF-HF and CFF91-HF predict binding energies in both sets of species with an MUE of less than 7.0 kcal/mol. Fourteen of the 16 methods tested predict binding energy for the charged species less accurately than for the neutral species, with some methods, such as AMBER and AMBER*, displaying a 4-fold increase in MUE. Two methods, AM1 and CVFF-CM4, predict binding energy for protonated species more accurately than for neutral species, with MUE differences between the two species of 1.0 and 2.0 kcal/mol, respectively.

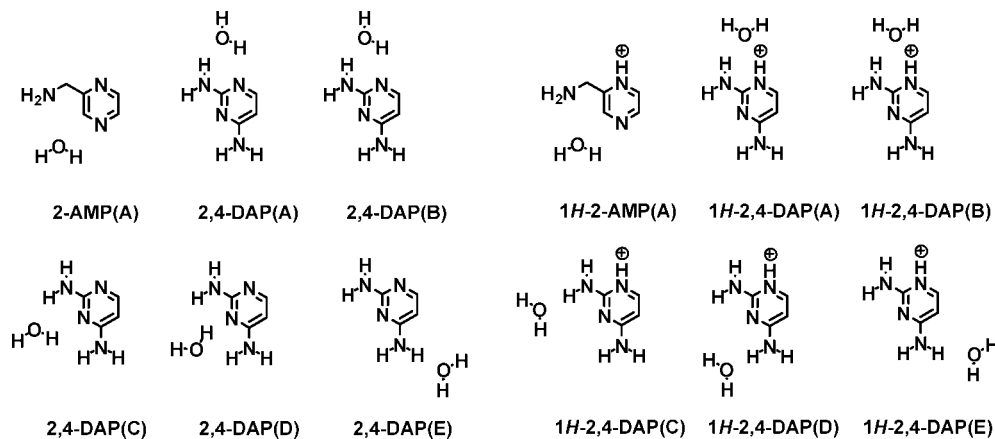


Figure 9. Molecular systems used in binding energy calculations. Position of the water molecule reflects the optimized complex.

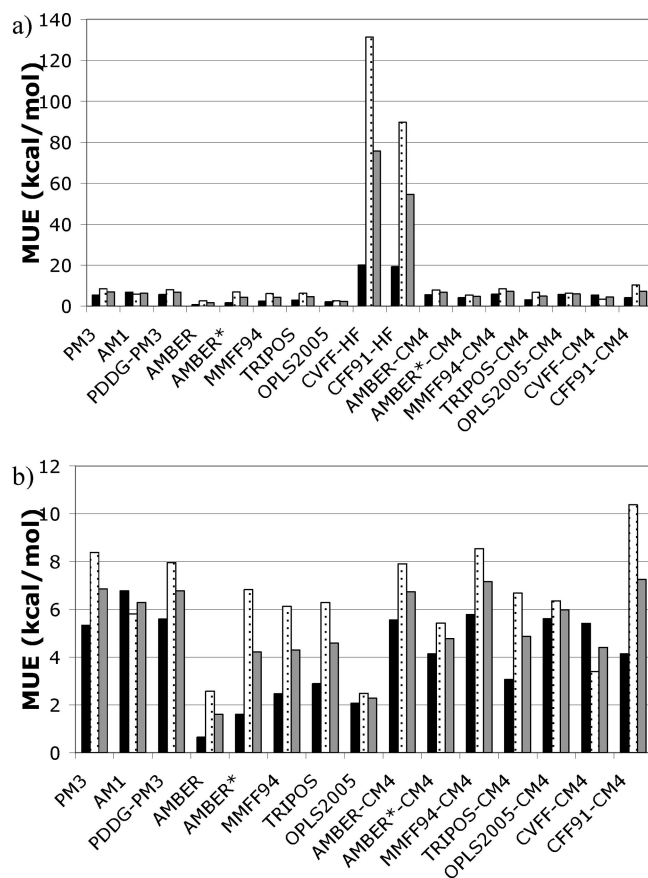


Figure 10. MUE in prediction of binding energy between 2-AMP, 1H-2-AMP, 2,4-DAP, and 1H-2,4-DAP and water molecules placed at hydrogen bonding locations denoted in the DHFR-MTX crystal structure. Panel a includes all methods tested, and panel b omits CVFF-HF and CFF91-HF for easier viewing (black, neutral species; dotted, charged species; gray, both sets of species).

The CVFF-HF and CFF91-HF force fields perform quite poorly, likely, because of the necessary substitution of CHPLPG charges for the force fields' default charges. Notably, the substitution of CM4 charges into the CVFF and CFF91 force fields again improves their modeling accuracy; however, this is not the case with the remaining MM force fields. Out of all the methods tested, the most

Table 5. Reduced Deviance ($D_{y,m}$) in Partial Charge for Force Field Validation For All Neutral and Charged Species in Figure 3

method	gas phase	aqueous phase	both phases
AMBER*	0.68	0.69	0.68
MMFF94	1.04	1.04	1.04
AMBER	1.05	1.07	1.06
OPLS2005	1.18	1.19	1.18

Table 6. Reduced Deviance ($D_{y,m}$) in Combined Geometry for Force Field Validation For All Neutral and Charged Species in Figure 3

method	gas phase	aqueous phase	both phases
MMFF94-CM4	0.73	0.79	0.76
MMFF94	0.80	0.85	0.82
OPLS2005-CM4	0.82	0.91	0.86
OPLS2005	0.95	0.88	0.92
AMBER-CM4	0.90	0.93	0.92
AMBER	1.03	0.97	1.00
AMBER*	1.38	1.32	1.34
AMBER*-CM4	1.39	1.35	1.37

accurate are the OPLS2005 and AMBER force fields, which predict binding energy for both sets of molecules to an MUE of 2.28 and 1.61 kcal/mol, respectively. In addition to its excellent accuracy, the OPLS2005 force field shows consistency between the neutral and charged species, calculating binding energy to an MUE of 2.07 and 2.49, respectively. The TRIPOS, MMFF94, and AMBER* force fields all predict binding energies about 2-fold less accurately, with MUEs of 4.59, 4.29, and 4.22 kcal/mol, respectively.

3.7. Validation of the AMBER, AMBER*, MMFF94, and OPLS2005 Force Fields. Tables 5 and 6 summarize our validation of the AMBER, AMBER*, MMFF94, and OPLS2005 force fields (as well as their MM-CM4 counterparts) in the gaseous and aqueous phases against both the molecules already included in the study, as well as an additional set of pharmacophore-containing molecules and their cations shown in Figure 3. In contrast to its relatively poor performance assigning partial charges to 2-AMP, 1H-2-AMP, 2,4-DAP, and 1H-2,4-DAP, the AMBER force field performs better in this case than OPLS2005, although the degree to which this increased performance is relevant is

Table 7. Partial Charge Distribution of Gas-Phase Neutral and Cationic MTX as Calculated by M05-2X/CM4 and MMFF94

MTX				MTX+			MTX				MTX+		
atom	CM4	MMFF94	residual	CM4	MMFF94	residual	atom	CM4	MMFF94	residual	CM4	MMFF94	residual
C1	0.44	0.72	0.28	0.51	0.77	0.26	C29	−0.12	0.09	0.21	−0.08	0.09	0.17
C2	0.29	0.41	0.12	0.27	0.41	0.14	H30	0.07	0.15	0.08	0.07	0.15	0.08
C3	0.07	0.31	0.24	0.09	0.31	0.22	H31	0.09	0.15	0.06	0.10	0.15	0.05
C4	0.29	0.62	0.33	0.35	0.67	0.32	C32	0.00	0.37	0.37	−0.01	0.37	0.38
C5	0.14	0.17	0.03	0.18	0.17	0.02	H33	0.06	0.00	0.06	0.07	0.00	0.07
C6	0.14	0.16	0.02	0.21	0.16	0.05	H34	0.06	0.00	0.06	0.06	0.00	0.06
H7	0.07	0.15	0.08	0.09	0.15	0.06	H35	0.06	0.00	0.06	0.06	0.00	0.06
N8	−0.59	−0.90	0.31	−0.52	−0.90	0.38	C36	0.36	0.54	0.19	0.33	0.54	0.22
H9	0.31	0.40	0.09	0.33	0.40	0.07	O37	−0.38	−0.57	0.19	−0.36	−0.57	0.22
H10	0.32	0.40	0.08	0.34	0.40	0.06	N38	−0.43	−0.73	0.30	−0.43	−0.73	0.30
N11	−0.61	−0.90	0.29	−0.55	−0.90	0.35	H39	0.26	0.37	0.11	0.26	0.37	0.11
H12	0.32	0.40	0.08	0.35	0.40	0.05	C40	0.11	0.36	0.26	0.11	0.36	0.25
H13	0.32	0.40	0.08	0.34	0.40	0.07	H41	0.07	0.00	0.07	0.07	0.00	0.07
N14	−0.44	−0.62	0.18	−0.41	−0.62	0.21	C42	−0.08	0.00	0.08	−0.07	0.00	0.07
N15	−0.44	−0.62	0.18	−0.40	−0.18	0.22	H43	0.07	0.00	0.07	0.07	0.00	0.07
N16	−0.32	−0.62	0.30	−0.31	−0.62	0.31	H44	0.07	0.00	0.07	0.07	0.00	0.07
N17	−0.29	−0.62	0.33	−0.28	−0.62	0.34	C45	−0.10	0.06	0.16	−0.10	0.06	0.16
C18	0.03	0.51	0.49	0.03	0.51	0.48	H46	0.09	0.00	0.09	0.09	0.00	0.09
H19	0.05	0.00	0.05	0.06	0.00	0.06	H47	0.09	0.00	0.09	0.09	0.00	0.09
H20	0.06	0.00	0.06	0.07	0.00	0.07	C48	0.29	0.66	0.37	0.29	0.66	0.37
N21	−0.33	−0.84	0.51	−0.34	−0.84	0.50	C49	0.29	0.66	0.37	0.29	0.66	0.37
C22	0.18	0.10	0.08	0.16	0.10	0.06	O50	−0.35	−0.57	0.22	−0.35	−0.57	0.22
C23	−0.11	−0.15	0.04	−0.12	−0.15	0.03	O51	−0.34	−0.57	0.23	−0.33	−0.57	0.24
C24	−0.11	−0.15	0.04	−0.10	−0.15	0.05	O52	−0.33	−0.65	0.33	−0.32	−0.65	0.33
C25	−0.05	−0.15	0.10	−0.04	−0.15	0.11	H53	0.32	0.50	0.18	0.33	0.50	0.18
H26	0.06	0.15	0.09	0.05	0.15	0.10	O54	−0.36	−0.65	0.29	−0.36	−0.65	0.29
C27	−0.02	−0.15	0.13	−0.01	−0.15	0.14	H55	0.32	0.5	0.18	0.32	0.50	0.18
H28	0.07	0.15	0.08	0.07	0.15	0.08	H56	N/A ^a	N/A ^a	N/A	0.33	0.46	0.13
							MUE						

representative fragments of methotrexate, as well for as an additional set of druglike molecules. Furthermore, the MMFF94 force field without charge substitution exhibits the second best geometric performance among the sixteen methods tested. However, in our binding energy studies, we find that excellent performance modeling geometries and charge distributions does not necessarily correlate directly to the prediction of energetics. Therefore, when the combined charge distribution, geometric, and energetic results are taken into account, we consider the MMFF94, AMBER/GAFF, AMBER*, and OPLS2005 force fields to be the most accurate and economical methods available for modeling small molecules containing nitrogen heterocycles and exocyclic amines. We expect these methods to be suitable for use in modeling more general nitrogen-containing small molecules as well as larger systems including the bis-methotrexate chemical inducer of dimerization, the protein–ligand complex, and the residues contained at the DHFR–DHFR interface.

Acknowledgment. We would like to thank David Ferguson, Casey Kelly, Benjamin Lynch, Nate Schultz, and Yuk Sham for their assistance during this study. B.R.W. was supported by a Chemistry and Biology NIH traineeship (T32 GM008700). This work was supported in part by the University of Minnesota Nanobiotechnology Initiative, the University of Minnesota Supercomputing Institute, and NSF grant CHE-0704974.

Supporting Information Available: Numbering system of 2-AMP, 1*H*-2-AMP, 2,4-DAP, 1*H*-2,4-DAP, MTX, and the MTX cation (Figures S1–S3), benchmark calculations of partial charge, bond length, and bond angle on 2-AMP, 1*H*-2-AMP, 2,4-DAP, and 1*H*-2,4-DAP utilizing CCSD/6–31+G(d,p) in the gas phase, as well as M05-2X/6–31+G(d,p) in both the gas and aqueous phases (Tables S1–S12), and cartesian coordinates of CCSD and M05-2X optimizations (Tables S13–S15). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Gohlke, H.; Klebe, G. *Agnew. Chem. Int. Ed.* **2002**, *41*, 2644.
- Hessler, G.; Klabunde, T. *ChemBioChem* **2002**, *3*, 928.
- Lugovsky, A. A.; Degterev, A. I.; Fahmy, A. F.; Zhou, P.; Gross, J. D. *J. Am. Chem. Soc.* **2002**, *124*, 1234.
- Perez, J. J.; Concho, F.; Llorens, O. *Curr. Med. Chem.* **2002**, *24*, 2209.
- Rao, G. S.; Bhatnagar, S.; Ahuja, V. *J. Biomol. Struct. Dyn.* **2002**, *20*, 31.
- Shahripour, A. B.; Plummer, M. S.; Lunny, E. A.; Albrecht, H. P.; Hays, S. J. *Bioorg. Med. Chem.* **2002**, *10*, 31.
- Anderson, A. *Chem. Biol.* **2003**, *10*, 787.
- Berlicki, L.; Kafarski, P. *Curr. Org. Chem.* **2005**, *9*, 1829.
- Ikejiri, M.; Bernardo, M. M.; Meroueh, S. O.; Brown, S.; Chang, M. *J. Org. Chem.* **2005**, *70*, 5709.
- Zhang, D. A. W.; Zhang, J. Z. H. *Int. J. Quantum Chem.* **2005**, *103*, 246.
- Armstrong, K. A.; Tidor, B.; Cheng, A. C. *J. Med. Chem.* **2006**, *49*, 2470.
- Ortiz, A. R.; Gomez-Puentas, P.; Leo-Macias, A.; Lopez-Romero, P.; Lopez-Vinas, E. *Curr. Top. Med. Chem.* **2006**, *6*, 41.
- Rao, G. S.; Ramachandran, M. V.; Bajaj, J. S. *J. Biomol. Struct. Dyn.* **2006**, *23*, 377.
- Ragno, R.; Simeoni, S.; Castellano, S.; Vicidomini, C.; Mai, A. *J. Med. Chem.* **2007**, *50*, 1241.
- Strockbine, B.; Rizzo, R. C. *Proteins: Struct., Funct., Genet.* **2007**, *67*, 630.
- Bowen, J. P.; Allinger, N. L. *Rev. Comput. Chem.* **1991**, *2*, 81.
- Dinur, U.; Hagler, A. T. *Rev. Comput. Chem.* **1991**, *2*, 99.
- Cornell, W. D.; Cieplak, P. *J. Am. Chem. Soc.* **1995**, *5179*.
- Petersson, I.; Liljefors, T. *Rev. Comput. Chem.* **1996**, *9*.
- MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D. *J. Phys. Chem. B* **1998**, *102*, 3586.
- Boyd, D. B.; Snoddy, J. D.; Lin, H. S. *J. Comput. Chem.* **1991**, *12*, 635.
- Jalaie, M.; Lipkowitz, K. *Rev. Comput. Chem.* **1999**, *14*.
- Khandelwal, A.; Lukacova, V.; Comez, D.; Kroll, D. M.; Raha, S. *J. Med. Chem.* **2005**, *48*, 5437.
- Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.
- Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.
- Mulholland, A. J. *Theor. Comput. Chem.* **2001**, *9*, 597.
- Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.
- Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
- Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Methods*; McGraw-Hill: New York, 1970.
- Dewar, M. J. S.; Ziegler, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
- Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221.
- Richards, N. J. *Molecular Orbital Calculations for Biological Systems*; Oxford University Press: New York, 1998.
- McKercher, S. R.; Lombardo, C. R.; Bobkov, A.; Jia, X.; Assa-Muut, N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 511.
- Elcock, A. H.; Sept, D.; McCammon, J. A. *J. Phys. Chem. B* **2001**, *105*, 1504.
- Prazulj, N.; Wigle, D. A.; Jurisica, I. *Bioinformatics* **2004**, *20*, 340.
- Gao, Y.; Wang, R.; Lai, L. *J. Mol. Model.* **2004**, *10*, 44.
- Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238.
- Zhao, H. X. *Curr. Med. Chem.* **2004**, *11*, 539.
- Hou, T.; Chen, K.; McLaughlin, W. A.; Lu, B.; Wang, W. *PLoS Comp. Biol.* **2006**, *2*, 46.
- Ababou, A.; van der Vaant, A.; Gogonea, V.; Merz, K. M. J. *Biophys. Chem.* **2007**, *125*, 221.
- Aslan, F. M.; Yu, Y.; Vajda, S.; Mohr, S. C.; Cantor, C. R. *J. Biotechnol.* **2007**, *128*, 213.

- (43) Carlson, J. C. T.; Kanter, A.; Thudappathy, G. R.; Cody, V.; Pineda, P. E. *J. Am. Chem. Soc.* **2003**, *125*, 1501.
- (44) Sutton, P. A.; Cody, V.; Smith, D. *J. Am. Chem. Soc.* **1986**, *108*, 4155.
- (45) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650.
- (46) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput-Aid Mol. Des.* **1995**, *9*, 87.
- (47) Cizek, J. *J. Chem. Phys.* **1966**, *45*, 4256.
- (48) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.
- (49) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.
- (50) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
- (51) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.
- (52) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
- (53) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 31.
- (54) Clark, M.; Cramer III, R. D.; van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982.
- (55) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (56) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
- (57) Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955.
- (58) Li, J.; Zhu, J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1998**, *102*, 1820.
- (59) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291.
- (60) Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *1000*, 16385.
- (61) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*.
- (62) Helgaker, T.; Gauss, J.; Jorgensen, P.; Olsen, J. *J. Chem. Phys.* **1997**, *106*, 6430.
- (63) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Phys. A* **2003**, *107*, 1384.
- (64) Wong, M. W.; Radom, L. *J. Phys. Chem. A* **1998**, *102*, 2237.
- (65) Byrd, E. F. C.; Sherrill, C. D.; Head-Gordon, M. *J. Phys. Chem. A* **2001**, *105*, 9736.
- (66) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (67) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.
- (68) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.
- (69) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.
- (70) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.
- (71) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 1202.
- (72) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
- (73) Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins-Winget, P. D.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM*, version 6.0; University of Minnesota: Minneapolis, MN, 2006.
- (74) Zhao, Y.; Truhlar, D. G. *MN-GFM Minnesota Gaussian Functional Module*, version 2.0.1; University of Minnesota: Minneapolis, MN, 2006.
- (75) Mulliken, R. S. *J. Chem. Phys.* **1935**, *3*, 564.
- (76) Lowdin, P. O. *J. Chem. Phys.* **1950**, *18*, 365.
- (77) Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428.
- (78) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
- (79) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
- (80) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
- (81) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
- (82) Franci, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. *J. Comput. Chem.* **1996**, *17*, 367.
- (83) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
- (84) Chirlian, L. E.; Franci, M. M. *J. Comput. Chem.* **1987**, *1987*, 894.
- (85) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601.
- (86) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Zhu, T.; Thompson, J. D.; Winget, P.; Lynch, B. J.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 7.1; University of Minnesota: Minneapolis, MN, 2004.
- (87) Chamberlin, A. C.; Pu, J.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D. *GAMESSPLUS*, version 4.7; University of Minnesota: Minneapolis, MN, 2005. Based on the General Atomic and Molecular Electronic Structure System (GAMESS) as described in Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupius, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
- (88) Stewart, J. J. P.; Zheng, J.; Rossi, I.; Hu, W.; Lynch, G. C.; Liu, Y.; Chuang, Y.; Pu, J.; Li, J.; Cramer, C. J.; Fast, P. L.; Truhlar, D. G. *MOPAC*, version 5.011mn; University of Minnesota: Minneapolis, MN, 2006.
- (89) Maple, J. R.; Hwang, M. J.; Stockfish, T. P.; Dinur, U.; Waldman, M. *J. Comput. Chem.* **1994**, *15*, 162.
- (90) Duan, Y.; Wu, C.; Choudhury, S.; Lee, M. C.; Xiong, G. *J. Comput. Chem.* **2003**, *24*, 1999.
- (91) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
- (92) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Morgan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.
- (93) Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*.
- (94) Brom, J. M.; Schmitz, B. J.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*.

- (95) Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, 108.
- (96) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, 113, 133.
- (97) Momany, F. J. *J. Phys. Chem.* **1978**, 85, 592.
- (98) Dinur, U.; Hagler, A. T. *J. Am. Chem. Soc.* **1989**, 111, 5149.
- (99) Hobza, P.; Kabelec, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1996**, 118, 1136.
- (100) McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, 33, 7747.
- (101) Chakravorty, S.; Reynolds, C. H. *J. Mol. Graphics Modell.* **1999**, 17, 315.
- (102) Halgren, T. A. *J. Comput. Chem.* **1999**, 20, 730.

CT8000766