JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Automated Parametrization of AMBER Force Field Terms from Vibrational Analysis with a Focus on Functionalizing Dinuclear Zinc(II) Scaffolds

Steven K. Burger,*[,†] Mike Lacasse,[†] Toon Verstraelen,[‡] Joel Drewry,[§] Patrick Gunning,[§] and Paul W. Ayers[†]

[†]Department of Chemistry and Chemical Biology, McMaster University, 1280 Main St. West, Hamilton, Ontario, Canada

[‡]Center for Molecular Modeling, Ghent University, 903 Technologiepark, B-9050 Zwijnaarde, Belgium

[§]Department of Chemistry, University of Toronto, 3359 Mississauga Road North, Mississauga, Ontario, Canada L5L 1C6

Supporting Information

ABSTRACT: A procedure for determining force constants that is independent of the internal redundant coordinate choice is presented. The procedure is based on solving each bond and angle term separately, using the Wilson B matrix. The method only requires a single *ab initio* frequency calculation at the minimum energy structure and is made available in the software "parafreq". The methodology is validated with a set of small molecules, by showing it can reproduce *ab initio* frequencies better than other methods such as taking the diagonal terms of the Hessian in internal coordinates or by using standard AMBER force fields. Finally, the utility of the method is demonstrated by parametrizing the dizinc scaffold of bis-dipicolylamine (BDPA) bound to phosphotyrosine, which is then functionalized into promising antitumor drug proteomimetics.

INTRODUCTION

Metal ions play a key role in the function of proteins. These ions can have a structural role, like zinc fingers,[1] which allow the protein to bind to various ligands, enabling its activity in cellular processes such as DNA repair, metabolism, and signaling. Alternatively, in metalloenzymes, the metal is directly involved in the catalytic mechanism.[2]

Metal ions also play important roles in metal organic complexes. In particular, there has been much interest in dizinc scaffolds,[3−5] which are often more efficient catalysts than monozinc complexes. These complexes can act to cleave nucleic acids.[6,7] They can also be used in the recognition of phosphates.[5,8] In particular, dizinc scaffolds have recently been shown to disrupt binding between proteins, as a proteomimetic of the phosphopeptide-binding domain of Stat3.[9−12] In this case, a functionalized bis-dipicolylamine (BDPA) binds to a phosphorylated tyrosine residue; the functional group can be tuned to recognize the surrounding residue sequences. Selecting the best functional group is challenging since there are myriad possibilities, many of which can be difficult to synthesize. With a well-parametrized system, one could guide the experiment by screening functional groups computationally and identifying the most promising compounds.

Even though metal complexes are ubiquitous in chemistry, there is no good, universal method for parametrizing such compounds. Part of the problem stems from the fact that most force fields are optimized to deal with organic systems, which easily retain their bonded structures. By contrast, metal system can change their coordination number and geometric structure[13] and generally do not have transferable force field parameters.

For enzyme systems, this problem is alleviated by the rigid structure of the protein. For metal complexes in solution, the problem is more acute, and new parameters need to be generated for each system.

There are a number of approaches for parametrizing metal complexes. One can ignore the bonded terms and approximate the potential energy surface using electrostatics and van der Waals terms.[14,15] This has the advantage of allowing for changes in coordination. It also reduces the number of unknown parameters that need to be determined. However, the paucity of parameters involving the metal center leads to unphysical rearrangements which give unrealistic coordination structures. The ion can even escape its coordination cage and dissociate from the ligand(s) entirely.

In bonded approaches, many more terms are introduced, specifically bond-stretching, angle-bending, and torsional coordinates that include the metal center. This generally fixes the coordination, which gives better agreement with the original crystal structure during a molecular dynamics run. With bonded terms, there is also the advantage of including the metal ions in the frequency analysis. An interesting comprise between the bonded and nonbonded is an approach referred to as the semibonded[16] method, which uses virtual atoms to keep zinc atoms in the correct orientation, although this approach is not widely used.

With a bonded approach, the force constants for the metal complexes can be obtained either by optimizing the force-field parameters directly[17−20] or by doing a frequency calculation.[21,22] Direct optimization is often very costly: the cost of (globally) optimizing the parameters grows exponentially as the number of parameters increases. Least-squares fitting,[23] Newton−Raphson optimization,[18] the simplex method,[24] and

Received: November 2, 2011
Published: January 4, 2012

   dx.doi.org/10.1021/ct2007742 | J. Chem. TheoryComput. 2012, 8, 554−562

genetic algorithms[25,26] have been used to approximate the terms by using a portion of the full parameter space. Specifically, in the case of metal sites, Hu and Ryde[27] have compared the frequency based method of Seminario[28] with the more exhaustive approach of Norrby and Liljefors.[18] They found that there is inevitable compromise between the computational and manual cost of parametrization and the resulting accuracy of the force field.

A frequency calculation allows one to identify bonded force constants with the diagonal elements of the Hessian in internal coordinates.[29−32] Unfortunately, this method is very sensitive to the choice of internal coordinate. Seminario and Bautista[28,33] give an example with the cyclic chon molecule, where the force constant changes by a factor of 2 depending on the choice of internal coordinate. To deal with the problem, Seminario proposed a method to obtain force constants by projection. Ryde[31] used this method for metalloproteins, while Seminario and Bautisuta et al.[33] applied it to small polypeptide chains. Lin and Wang[34] and Peters et al.[35] have used the method to derive parameters for systems containing zinc atoms.

In this article, we solve for the force constant of each internal coordinate separately.[36] This avoids the problem of coordinate choice and closely reproduces the molecular Hessian matrix. A python code, available on our Web site, allows researchers to use this method for their own parametrization.

METHODOLOGY

Our aim is to work with the AMBER[37] force field, defined by the energy function

$$E = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}(1 + \cos[n\varphi - \delta]) + \sum_{i,j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}} \quad (1)$$

which consists of terms for bonds, angles, dihedrals, van der Waals, and electrostatics. For most systems, these terms are derived with the aid of the software Antechamber[38] using either an AMBER amino acid force field or the general AMBER force field (GAFF).[39] Antechamber assigns terms based on connectivity, and generally the parameters work well for organic systems. However, for metal complexes and for systems with features not represented in the standard library set, many terms are either unassigned or are inaccurate (and need to be redone).

QM Frequency Calculation for Bonds and Angles. To obtain unassigned bond and angle terms, a sequence of energy calculations needs to be done at both the QM and MM levels. The difference between the potential energy surfaces can then be fit to determine the harmonic constants. However, this is generally a computationally expensive procedure. More efficiently, these terms can be determined from a single frequency calculation with the minimum energy state structure. As with a relaxed scan, the vibrational frequencies can be computed at both the QM level and the MM level, so that the effects already present in the force field can be removed. When the difference between the Hessians in Cartesian coordinates is written as $\mathbf{H}_x = \mathbf{H}_x^{\text{QM}} - \mathbf{H}_x^{\text{MM}}$, the force constants for the internal coordinates can be derived using the Wilson B matrix[40]

$$B_{ij} = \frac{\partial q_i}{\partial x_j} \quad (2)$$

where q_i are the internal coordinates and x_j are the Cartesian coordinates. The Hessian in internal coordinates is

$$\mathbf{H}_q = (\mathbf{B}^T)^+(\mathbf{H}_x - \mathbf{K})\mathbf{B}^+ \quad (3)$$

where $\mathbf{B}^+$ is the Moore−Penrose pseudoinverse of the Wilson B matrix, $K_{jk} = \sum_i g_i^q(\partial^2 q_i)/(\partial x_j \partial x_i)$, and g_i^q is an element of the internal coordinate gradient. It is important to evaluate $\mathbf{K}$ because, even though we evaluate the Hessian at the optimal geometry, $\mathbf{g}^q$ may not be fully zero. Nevertheless, this does add a small coordinate-dependent correction to the force constant. Its inclusion into the algorithm was not found to significantly affect the results.

If the internal coordinates fully describe the system, and $\mathbf{H}_q$ is a diagonal matrix, then the force constant parameters are exact and then can be obtained from the diagonal. When this is not the case, two problems can emerge from eq 3: (1) The transformation is sensitive to the choice of internal coordinates. (2) Significant cross terms appear in the Hessian matrix which couple different internal coordinates; these terms can be accounted for directly in the force field but do not appear in eq 1.

Rather than solving for the full Hessian in internal coordinates, we propose instead to solve for each internal coordinate separately. In this case, the matrix $\mathbf{B}$ is simply a vector with the only nonzero terms being the atoms involved in the internal coordinate. Following ref 36, we have:

1. For each bond between atoms m and n, the $\mathbf{B}$ matrix is defined as

$$B_i = \frac{\partial q_b}{\partial a_i} = \xi_{amn} v_i,\ a \in \{m, n\},\ i = x, y, z \quad (4)$$

where $\mathbf{v} = \mathbf{m} - \mathbf{n}/\|\mathbf{m} - \mathbf{n}\|$ and $\xi_{amn} = (\delta_{am} - \delta_{an})$ is the difference between the two Dirac delta functions centered at the atom a. The value ξ_{amn} is only nonzero when a is equal to either m or n.

2. For an angle between atoms m, o, and n

$$B_i = \frac{\partial q_b}{\partial a_i} = \xi_{amo}\frac{[\mathbf{u} \times (\mathbf{u} \times \mathbf{v})]_i}{\|\mathbf{m} - \mathbf{o}\|} + \xi_{ano}\frac{[(\mathbf{u} \times \mathbf{v}) \times \mathbf{v}]_i}{\|\mathbf{n} - \mathbf{o}\|},\ a \in \{m, n\},\ i = x, y, z \quad (5)$$

where $\mathbf{u} = \mathbf{m} - \mathbf{o}/\|\mathbf{m} - \mathbf{o}\|$ and $\mathbf{v} = \mathbf{n} - \mathbf{o}/\|\mathbf{n} - \mathbf{o}\|$. The dihedral terms can be written in a similar fashion. Using these single coordinate transformations, the Cartesian Hessian can be mapped to the force constant of interest with eq 3. While eq 3 is normally used as a matrix equation to map the entire Cartesian Hessian into the internal coordinate space, in this case, each $\mathbf{B}$ matrix is only a column vector, and so $\mathbf{H}_q$ is a 1×1 matrix. The single value of the matrix is the force constant.

dx.doi.org/10.1021/ct2007742 | J. Chem. TheoryComput. 2012, 8, 554−562

Since the second derivative terms are not independent, the force constants are determined iteratively starting with an initial guess. The electrostatics and van der Waals terms are included in the MM Hessian and are not allowed to change. Force constants related to internal coordinates could be held fixed. This is usually done with dihedrals which are better determined by scanning over the potential energy surface and, with bonds and angles that are considered to be reliable, are not expected to deviate significantly from GAFF values. An example would be the C−C bonds within aromatic rings. The full algorithm is as follows:

Algorithm for Determining Force Constants

1. Set initial values for the force constants $\sigma = \sigma_0$.
2. Compute the *ab initio* frequencies and the Cartesian Hessian, $\mathbf{H}_x^{QM}$.
3. Compute the AMBER Hessian matrix, $\mathbf{H}_x^{MM}$.
4. Set $\mathbf{H}_x = \mathbf{H}_x^{QM} - \mathbf{H}_x^{MM}$.
5. Compute $\mathbf{K}$.
6. For each bond and angle i:
   a. Compute $\mathbf{B}_i$.
   b. $\mathbf{B}_i^+ = \text{pseudoinverse}(\mathbf{B}_i)$
   c. Solve $\Delta\sigma_i = (\mathbf{B}_i^T)^+(\mathbf{H}_x - \mathbf{K})\mathbf{B}_i^+/2$
7. Update the parameter files: $\sigma = \sigma + \Delta\sigma/2$
8. If $\|\Delta\sigma\| < 1.0$, then STOP; else, GOTO 2.

In step 6, we damp the step by a factor of 2 to prevent steps that are too large. The method is stopped when the change in force constants in units of kcal/mol/A^2 and kcal/mol/radian^2 drops below a threshold, here 1. The only other significant parameter which requires attention is the cutoff for the singular values when inverting the Wilson **B** matrix. This is addressed in the Computational Details section.

Dihedral Terms. While the dihedral terms could be fit to vibrational frequencies, the harmonic approximation is particularly poor for the torsional degrees of freedom. Instead, the terms are usually derived from a relaxed PES scan. Specifically, a scan of the dihedral angle is performed both with a quantum mechanics (QM) method and with a molecular mechanics (MM) force field in which the dihedral term has been set to zero. The difference in energy between the two scans is then fit to a Fourier expansion:

$$\sum_{i=1}^{M} \frac{V_i}{2}(1 + \cos[in\varphi - \delta]) \tag{6}$$

where M is the number of terms that are used, V_i represents the Fourier coefficients, δ is the phase shift, and n is the periodicity. While any number of terms may be used in the expansion, to guard against overfitting, it is wise to use a minimal number of terms. Generally, one sets M to truncate terms in the Fourier series whose coefficients are sufficiently small. We chose to expand eq 6 until the coefficients were less than 0.1. The scans were done with fixed increments, giving a least-squares equation

$$\begin{pmatrix} V_{QM}(\varphi_1) - V_{MM}(\varphi_1) \\ \vdots \\ V_{QM}(\varphi_N) - V_{MM}(\varphi_N) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} (1 + \cos[n\varphi_1 - \delta]) & \cdots & (1 + \cos[Mn\varphi_1 - \delta]) \\ \vdots & \ddots & \vdots \\ (1 + \cos[n\varphi_N - \delta]) & \cdots & (1 + \cos[Mn\varphi_N - \delta]) \end{pmatrix}\begin{pmatrix} V_1 \\ \vdots \\ V_M \end{pmatrix} \tag{7}$$

to solve for the Fourier coefficients.

COMPUTATIONAL DETAILS

The method was first tested with several small molecule systems: hydrogen peroxide, biphenyl, bicyclo[2.2.2]octane, 3-nitropyridine, and toluenesulfonic acid. These structures were built with Gaussview[41] and then optimized with Gaussian 09[42] (g09) using HF/6-31G(d). Charges were obtained with electrostatic potential (ESP) fitting using the pop=CHELPG[43] keyword.

In addition, we parametrized the metal-containing system bis-dipicolylamine (BDPA) which consists of two bound zinc ions and a phosphotyrosine. This was split into two parts to be parametrized separately: (1) the biphenyl ring (Figure 1) and

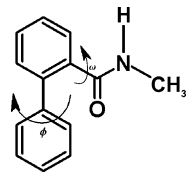

Figure 1. The two torsional angles adjusted to match the QM values. The angle ϕ was used to generate the profile in Figure 4.

(2) the dizinc BDPA scaffold bound to phosphotyrosine (Figure 2). For the BDPA scaffold, the bonded model was used with a coordinate bond added between the five neighboring atoms of each Zn^{2+} ion.

For all systems, Cartesian Hessians and vibrational frequencies were determined with the FREQ keyword and the "freqchk" utility provided with g09. These values were then scaled with the appropriate scaling factor,[44] 0.8953, for Hartree−Fock. The MM frequencies were calculated with the NAB[45] (Nucleic Acid Builder) molecular manipulation language. This has a C-like interface to AMBER. The program LEAP, which is included in AmberTools 1.4,[37] was used to generate the parameter files. A short code was written to read in the parameter files and coordinates and then call the function mme2 to compute the MM Hessian.

The MM Hessian matrix was subtracted from the QM one before determining the force constants. Force constants were determined using both the traditional approach of taking the diagonal elements of the Hessian in internal coordinates and the approach proposed here, where each force constant is

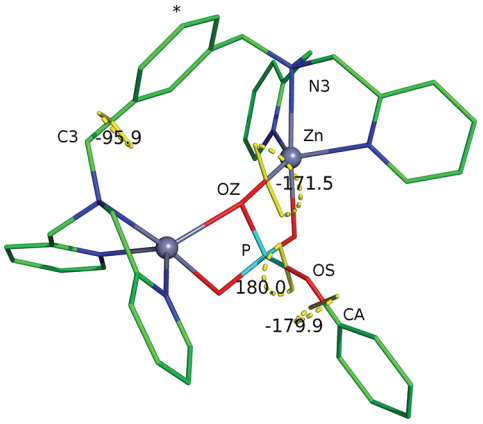

Figure 2. BDPA dihedrals scanned to determine the parameters. R groups were later placed at the starred position on the phenyl ring to derivatize the scaffold.

determined separately. In both approaches, the same basic iterative *Algorithm for Determining Force Constants* was used. When several bonds or angles had the same atom types, the average value of the force constant over all equivalent bonds/angles in the molecule was used. Equilibrium values for bonds and angles were taken from the QM structure.

The algorithm was implemented in Python with the aid of the package MolMod.[46] MolMod is a Python package that provides modules for internal coordinate transformations and derivatives using Gaussian formatted checkpoint files. This code was named *parafreq*, and it can be found on the Ayers group Web site.[47] The code can be used to determine force field parameters for bonds, angles, and dihedrals. While including dihedral terms gives more accurate frequencies, the resulting dihedral barriers can deviate significantly from barriers determined by a potential energy surface scan.

In the dizinc system, the improper dihedrals, atom types, dihedrals, and terms related to the organic part of the system were obtained with the program Antechamber[38] using GAFF and AMBER ff99SB. Additional terms related to the phosphotyrosine were taken from ref 48. For the nonorganic part of the dizinc system, the dihedrals were set to zero. Certain critical dihedrals that were important for the overall motion of the BDPA relative to the phosphotyrosine and dihedrals related to the motion of the biphenyl group were determined with a one-dimensional dihedral scan in g09 using the MODRED keyword. These dihedrals are shown in Figures 1 and 2. The FF terms were then fit to the error between the QM and MM energy profiles using eq 6. These critical dihedrals acted as catch-all terms to get the overall motion of the complex correct.

For the biphenyl on the BDPA, a methylated amino group (NME cap) was used to cap the R group (Figure 3) in the initial parametrization. The phosphotyrosine similarly was capped with an NME and an ACE group at either end. The charges on these capped groups were fixed to the standard AMBER values using restrained electrostatic potential[49] (RESP) fitting, with a 0.0005 au hyperbolic restraint on the heavy atoms.

These derivatized structures were then placed within the dizinc scaffold bonded to the pY group. The final sequence was created with LEAP using ACE and NME caps for the ends. The polypeptide was then minimized using steepest descent, followed by conjugate gradient. This was followed by heating the system for 100 ps to 300 K. Finally, a 4 ns production run

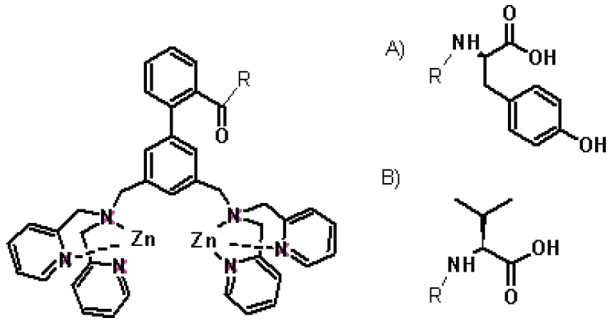

Figure 3. Two examples of amino acid functionalization of the BDPA receptor.

of MD with implicit solvent using a Berendsen thermostat and no cutoffs was performed.

RESULTS

Hydrogen Peroxide. As the simplest realistic molecule with bonds, angles, and dihedrals, hydrogen peroxide is an ideal test case for examining how well various force field parametrization techniques reproduce the vibrational frequencies from QM calculations. Table 1 compares the frequencies of the

Table 1. Frequencies (cm^{-1}) for Hydrogen Peroxide Using the Two Possible Methods of Determining Force Constants

method	HF/6-31G(d)	diagonal terms	single projection
	3872.8	3913.0	3870.4
	3871.0	3902.6	3859.5
	1546.4	1288.4	1571.3
	1412.4	1184.8	1479.6
	1088.26	979.7	1078.0
	376.3	233.04	233.3
error	**0**	**809.33**	**259.26**

single-projection method introduced here to the frequencies obtained when the force constants are taken from the diagonal terms of the Hessian in internal coordinates. The frequencies shown are not scaled. The dihedral is left as a constant, and the internal force constants for, H−O, O−O, and H−O−O are allowed to vary.

The error in the frequencies is about a factor of 3 lower with the single projection method after three iterations of MM frequency calculations. Interestingly, the reduced error in the frequencies does not mean that the Hessian matrix itself is more accurate. In fact, if we compare the entrywise 2-norm, $\|A\|_2 = (\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2)^{1/2}$, of the difference between the parametrized Hessian matrix and the exact QM Hessian (both in Cartesian coordinates), using the diagonal terms is found to lead to a better approximation. Specifically, $\|\mathbf{H}\|_2/N = 0.22$ for the single-projection approach and 0.15 for the Hessian-diagonal approach.

Small Molecular Systems. The unscaled vibrational frequencies of the small molecules listed in Table 2 are compared to the "exact" HF/6-31G(d) results from Gaussian. The atom types, dihedrals, and improper terms were determined using GAFF; the remaining terms, which can be found in Tables S1−S4, were determined from the frequency analysis. From Table 2, it is clear that determining the force constants one at a time by the single projection technique gives the best results. The error obtained when using the GAFF

dx.doi.org/10.1021/ct2007742 | J. Chem. TheoryComput. 2012, 8, 554−562

Table 2. Norm of the Error in the Frequencies for Four Small Molecules

	single	diagonal	GAFF	seminario
biphenyl	62.2	521.7	129.98	80.71
bicyclo[2.2.2]octane	68.0	309.5	73.5	129.43
3-nitropyridine	78.7	115.6	156.5	128.70
toluenesulfonic acid	46.7	104.5	158.0	108.36

parameters is about twice as large, while the error from Seminario's method is about 1.5 times larger. When the force constants are determined from the diagonal of the force-constant matrix in internal coordinates, the vibrational frequencies have even larger errors. The error is particularly large for the biphenyl and the bicyclo[2.2.2]octane systems. This is probably due to linear dependencies between the internal coordinates: the Wilson B matrix had a large condition number in both cases.

Table 3. The Matrix Norm of the Difference to the QM Cartesian Hessian

	single	diagonal	GAFF	seminario
biphenyl	0.65	3.12	0.93	1.11
bicyclo[2.2.2]octane	0.23	0.69	0.50	0.38
3-nitropyridine	0.70	0.88	0.87	0.76
toluenesulfonic acid	0.50	0.66	1.12	0.60

Table 3 gives the norm difference of the QM Cartesian Hessian. The results largely mirror the relative errors in the frequencies shown in Table 2.

Functionalized Bis-dipicolylamine Dizinc Complex Bound to Phosphotyrsine. For the biphenyl ring, a frequency calculation was done to determine all bond and angle terms in the FF. These terms were similar to the GAFF values and were not used in the subsequent analysis. However, it is instructive to compare the exact QM frequency spectrum with the spectrum of the other two methods: taking the diagonal terms and determining each term separately. As seen in Figure 4, the agreement is very good for the algorithm for the separately determined spectrum.

For the dihedral angles of the biphenyl ring, we focused on the ones shown in Figure 1, since they affect the conformational space sampled most significantly. Comparing the QM torsional profile to the AMBER profile revealed differences on the order of 10−20 kcal/mol for two of the most important dihedral angles. This large difference was due, in part, to additional improper dihedrals imposed on the amide part of the molecule by Antechamber. It was also partly due to the effects of hyperconjugation, which are not accurately modeled by GAFF.

Once the improper and dihedral terms corresponding to these rotations were removed, the dihedral data were refit based on a Fourier expansion of the energy expression. For the dihedral corresponding to a rotation about the C−C bond between the biphenyl rings, the results of the match are shown in Figure 5.

For the dizinc scaffold, a similar procedure was performed with the dihedrals shown in Figure 2. These parameters could not be correctly determined with a frequency analysis, and the force constants were significantly different from those obtained with an energy scan. With the dihedral fixed, the value obtained with the scan *parafreq* was used to get the frequencies for the dizinc scaffold bound to the phosphorylated tyrosine. The spectrum is shown in Figure 6. The spectra match up well except for high-frequency stretches around 2400 cm^{-1} and 3000 cm^{-1}. These correspond to the C−C and C−N bonds in the phenyl rings and to C−H bonds, respectively, and are taken from AMBER ff99SB. Notably, some of the frequencies are negative, which is the result of fixing terms to the AMBER force field.

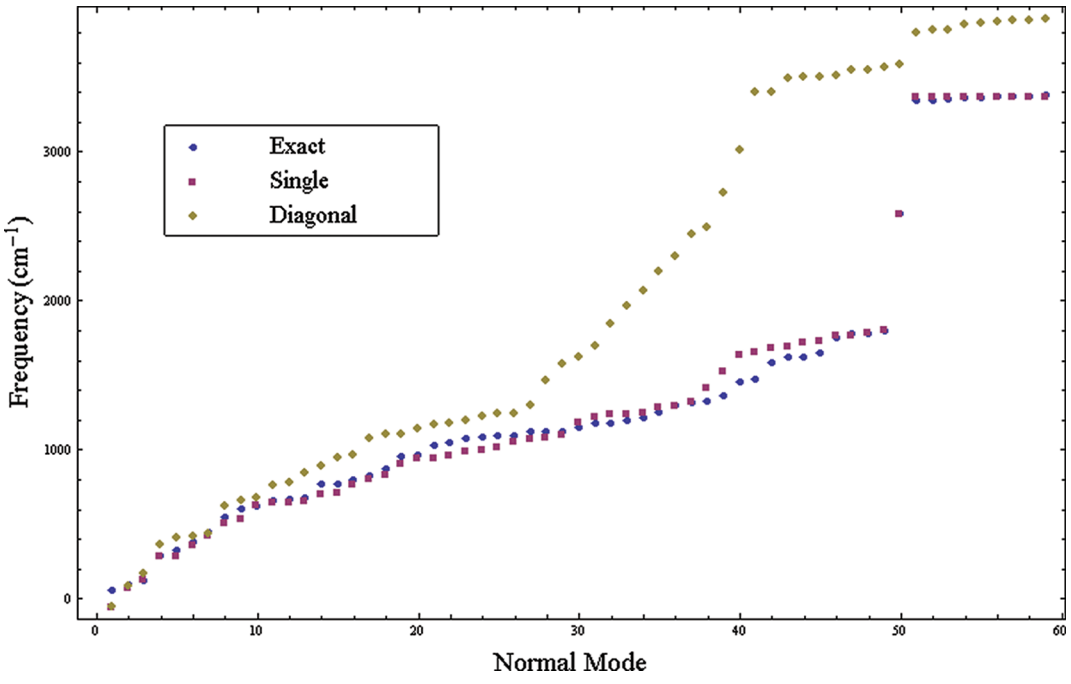

Figure 4. Frequency spectrum for the biphenyl group using HF/6-31G*. "Exact" is the QM spectrum.

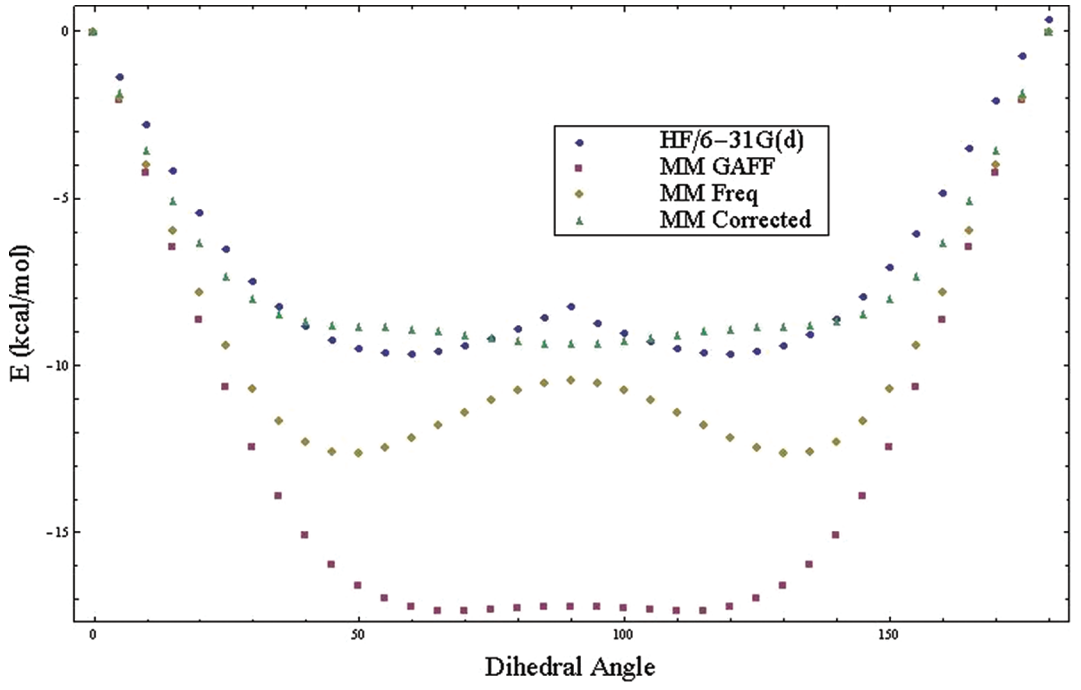

Figure 5. Torsional profile for the dihedral of the biphenyl system (Figure 1) using HF/6-31G(d), compared to using MM approaches. The MM corrected profile used three terms in the Fourier expansion.

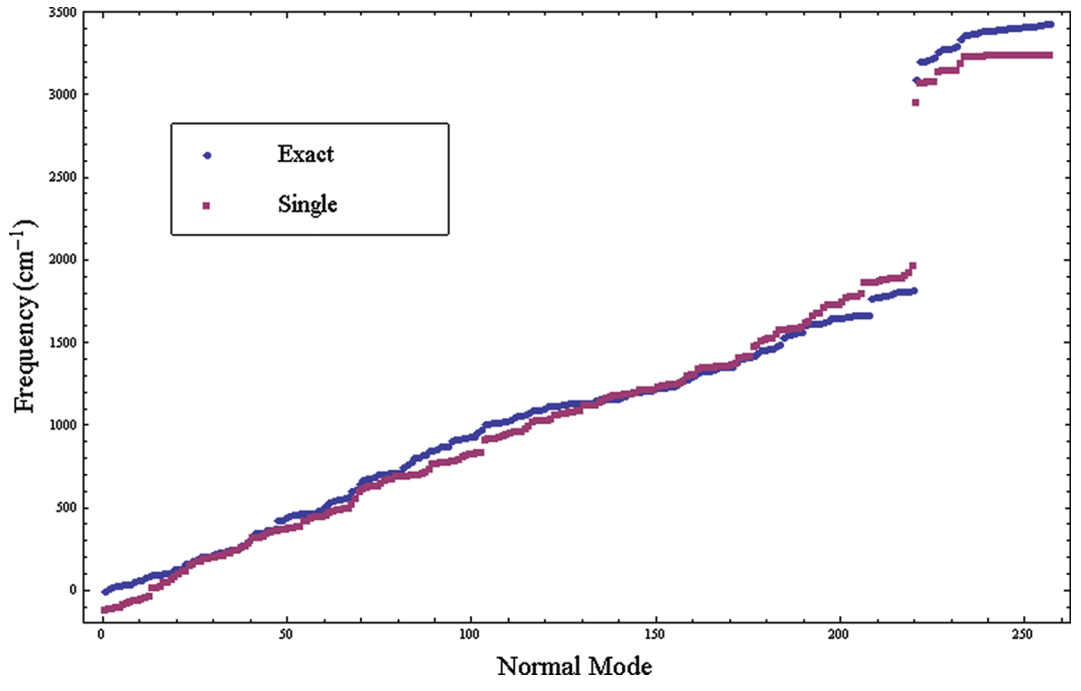

Figure 6. Frequency spectrum of the dizinc scaffold of the BDPA−phosphotyrosine complex shown in Figure 4. Exact is the QM spectrum.

The final structure was derivatized with two groups (Figure 3). These derivatized structures could then be used as custom residues in short polypeptide sequences. To test the stability of the structure when the parameters were derived in this way, short MD runs were performed on the sequence ACE−pYLKTK−NME, shown in Figure 7 for derivative B (Figure 3). An analysis of the trajectory (Figure 8) shows that the peptide samples a wide range of the conformations for the first few nanoseconds of the trajectory and then settles into an energetically stable structure.

The coordination does not change during the MD run, and the dizinc core remains similar to the minimum energy conformation. This points to the benefits and shortcomings of the method. Using an implicit solvent model with Gaussian, the enthalpic cost of adding a water into the coordination of the zinc in BDPA is 3−4 kcal/mol. A similar enthalpic cost was associated with a nonsymmetric structure, in which one site was tetrahedral and the other was pentacoordinated. Neither of these structures can be correctly accounted for with a bonded model derived through frequency analysis. Nevertheless, the

dx.doi.org/10.1021/ct2007742 | J. Chem. TheoryComput. 2012, 8, 554−562

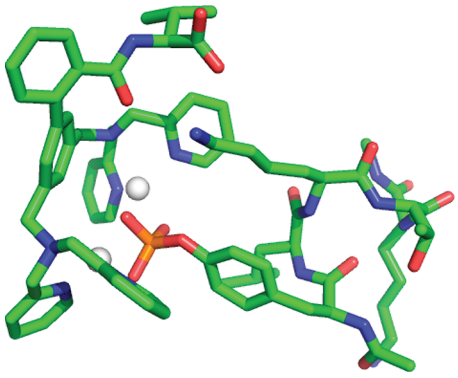

Figure 7. Lowest potential energy structure of ACE−pYLKTK−NME bound to BDPA functionalized with Val.

basic structure is maintained in a way that a nonbonded model could not achieve.

CONCLUSION

In this paper, we have presented a method for parametrizing the AMBER force field that is largely independent of the choice of coordinates. The method requires only a frequency calculation and is able to determine unknown bond or angle force constants from an existing partial parametrization. A corresponding Python software package called *parafreq* was developed that allows researchers to easily determine force constants for any system.

The method is based on solving for each internal coordinate force constant separately using the Wilson B Matrix. A test set of small molecules was used to show that the method gives significantly smaller errors in the frequencies compared to using parameters from GAFF or when using the diagonal elements from the Hessian in internal coordinates.

We further demonstrated that the method is applicable to large systems of interest, by parametrizing a biphenyl functionalized dizinc scaffold. For the biphenyl ring, we showed that one cannot rely upon GAFF parameters for the dihedrals, especially when hyperconjugation is involved. For the rest of the system, the frequency spectrum of the MM parametrized system was shown to match up well with the QM results, with significant deviations only in the organic region where the force constants were borrowed from the AMBER force field. The parametrization was verified with a molecular dynamics run in which the residue is incorporated into a short polypeptide strand. The MD settled in a stable structure with the correct coordination geometry.

There are some drawbacks, as there are with other methods based on harmonic frequencies, that still need to be addressed: (1) Dihedrals need to be determined by constrained optimizations, and (2) with a bonded model there is no easy way to account for changes in coordination. However, we have shown that with *parafreq*, molecular systems with unknown harmonic terms can be easily determined on the basis of a single frequency calculation, and that the frequency spectrum can be more accurately determined than other methods such as

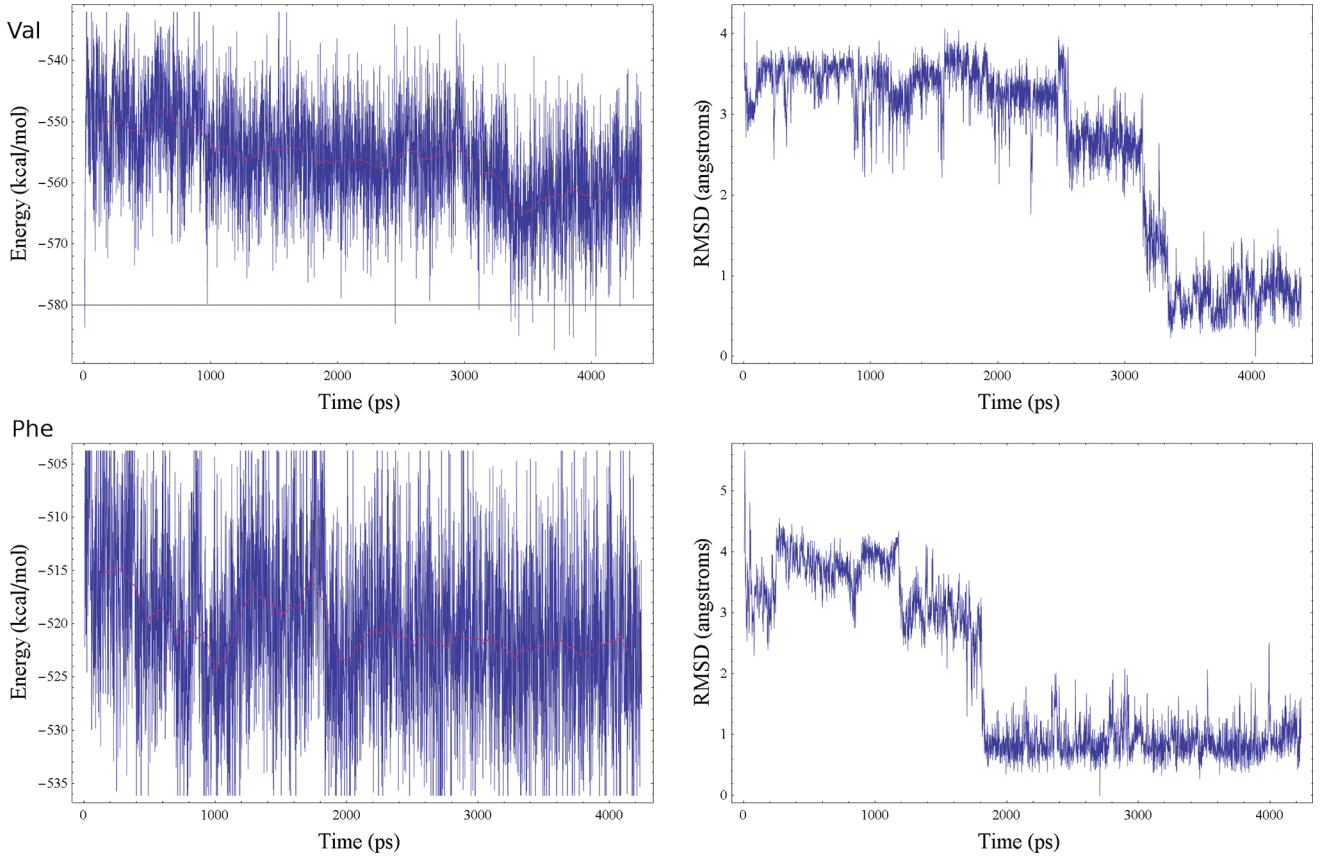

Figure 8. Potential energy and RMSD (relative to the lowest energy structure) of the polypeptide backbone of the Stat3 sequence ACE−pYLKTK−NME bound to BDPA. The red line is the 100 ps running average.

dx.doi.org/10.1021/ct2007742 | J. Chem. TheoryComput. 2012, 8, 554−562

using the diagonal elements from the internal coordinate Hessian, using GAFF parameters, or using Seminario's method.

■ ASSOCIATED CONTENT

Supporting Information

Bond and angle stretching parameters for biphenyl, bicyclo[2.2.2]octane, 3-nitropyridine, and toluenesulfonic acid. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: burgers@mcmaster.ca.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the chemistry department of McMaster University, NSERC, and the Canada Research Chairs. T.V. would like to thank Veronique Van Speybroek, Dimitri Van Neck, and Michel Waroquier for their continued encouragement and support and the European Research Council for funding under the European Community's Seventh Framework Programme (FP7(2007-2013) ERC grant agreement number 240483).

■ REFERENCES

(1) Krishna, S. S.; Majumdar, I.; Grishin, N. V. Structural classification of zinc fingers. *Nucleic Acids Res.* **2003**, *31* (2), 532−550.

(2) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **1999**, *68*, 779−819.

(3) Yashiro, M.; Kaneiwa, H.; Onaka, K.; Komiyama, M. Dinuclear Zn2+ complexes in the hydrolysis of the phosphodiester linkage in a diribonucleoside monophosphate diester. *Dalton Trans.* **2004**, *4*, 605−610.

(4) Yashiro, M.; Ishikubo, A.; Komiyama, M. Preparation and Study of Dinuclear Zinc(Ii) Complex for the Efficient Hydrolysis of the Phosphodiester Linkage in a Diribonucleotide. *J. Chem. Soc., Chem. Commun.* **1995**, *17*, 1793−1794.

(5) Kinoshita, E.; Takahashi, M.; Takeda, H.; Shiro, M.; Koike, T. Recognition of phosphate monoester dianion by an alkoxide-bridged dinuclear zinc(II) complex. *Dalton Trans.* **2004**, *8*, 1189−1193.

(6) Gao, H.; Ke, Z.; DeYonker, N. J.; Wang, J.; Xu, H.; Mao, Z.-W.; Phillips, D. L.; Zhao, C. Dinuclear Zn(II) Complex Catalyzed Phosphodiester Cleavage Proceeds via a Concerted Mechanism: A Density Functional Theory Study. *J. Am. Chem. Soc.* **2011**, *133* (9), 2904−2915.

(7) Yang, M. Y.; Iranzo, O.; Richard, J. P.; Morrow, J. R. Solvent deuterium isotope effects on phosphodiester cleavage catalyzed by an extraordinarily active Zn(II) complex. *J. Am. Chem. Soc.* **2005**, *127* (4), 1064−1065.

(8) Ojida, A.; Mito-oka, Y.; Inoue, M.; Hamachi, I. First artificial receptors and chemosensors toward phosphorylated peptide in aqueous solution. *J. Am. Chem. Soc.* **2002**, *124* (22), 6256−6258.

(9) Drewry, J. A.; Gunning, P. T. Recent advances in biosensory and medicinal therapeutic applications of zinc(II) and copper(II) coordination complexes. *Coord. Chem. Rev.* **2011**, *255* (3−4), 459−472.

(10) Drewry, J. A.; Fletcher, S.; Yue, P.; Marushchak, D.; Zhao, W.; Sharmeen, S.; Zhang, X.; Schimmer, A. D.; Gradinaru, C.; Turkson, J.; Gunning, P. T. Coordination complex SH2 domain proteomimetics: an alternative approach to disrupting oncogenic protein-protein interactions. *Chem. Commun.* **2010**, *46* (6), 892−894.

(11) Fletcher, S.; Drewry, J. A.; Shahani, V. M.; Page, B. D. G.; Gunning, P. T. Molecular disruption of oncogenic signal transducer and activator of transcription 3 (STAT3) protein. *Biochem. Cell Biol.* **2009**, *87* (6), 825−833.

(12) Drewry, J. A.; Fletcher, S.; Hassan, H.; Gunning, P. T. Novel asymmetrically functionalized bis-dipicolylamine metal complexes: peripheral decoration of a potent anion recognition scaffold. *Org. Biomol. Chem.* **2009**, *7* (24), 5074−5077.

(13) Kowall, T.; Foglia, F.; Helm, L.; Merbach, A. E. Molecular-Dynamics Simulation Study of Lanthanide Ions Ln(3+) in Aqueous-Solution Including Water Polarization - Change in Coordination-Number from 9 to 8 Along the Series. *J. Am. Chem. Soc.* **1995**, *117* (13), 3790−3799.

(14) Stote, R. H.; Karplus, M. Zinc-Binding in Proteins and Solution - a Simple but Accurate Nonbonded Representation. *Proteins: Struct., Funct., Genet.* **1995**, *23* (1), 12−31.

(15) Vedani, A.; Huhta, D. W. A New Force-Field for Modeling Metalloproteins. *J. Am. Chem. Soc.* **1990**, *112* (12), 4759−4767.

(16) Pang, Y. P. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins: Struct., Funct., Genet.* **2001**, *45* (3), 183−189.

(17) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. An automatic method to generate force-field parameters for hetero-compounds. *Acta Crystallogr., Sect. D* **2003**, *59*, 274−289.

(18) Norrby, P. O.; Liljefors, T. Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data. *J. Comput. Chem.* **1998**, *19* (10), 1146−1166.

(19) Norrby, P. O.; Brandt, P. Deriving force field parameters for coordination complexes. *Coord. Chem. Rev.* **2001**, *212*, 79−109.

(20) Norrby, P. O. Selectivity in asymmetric synthesis from QM-guided molecular mechanics. *THEOCHEM* **2000**, *506*, 9−16.

(21) Dasgupta, S.; Yamasaki, T.; Goddard, W. A. The Hessian biased singular value decomposition method for optimization and analysis of force fields. *J. Chem. Phys.* **1996**, *104* (8), 2898−2920.

(22) Vaiana, A. C.; Cournia, Z.; Costescu, I. B.; Smith, J. C. AFMM: A molecular mechanics force field vibrational parametrization program. *Comput. Phys. Commun.* **2005**, *167* (1), 34−42.

(23) Liang, G. Y.; Fox, P. C.; Bowen, J. P. Parameter analysis and refinement toolkit system and its application in MM3 parameterization for phosphine and its derivatives. *J. Comput. Chem.* **1996**, *17* (8), 940−953.

(24) Faller, R.; Schmitz, H.; Biermann, O.; Muller-Plathe, F. Automatic parameterization of force fields for liquids by simplex optimization. *J. Comput. Chem.* **1999**, *20* (10), 1009−1017.

(25) Wang, J. M.; Kollman, P. A. Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comput. Chem.* **2001**, *22* (12), 1219−1228.

(26) Tafipolsky, M.; Schmid, R. Systematic First Principles Parameterization of Force Fields for Metal-Organic Frameworks using a Genetic Algorithm Approach. *J. Phys. Chem. B* **2009**, *113* (5), 1341−1352.

(27) Hu, L. H.; Ryde, U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *J. Chem. Theory Comput.* **2011**, *7* (8), 2452−2463.

(28) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* **1996**, *60* (7), 1271−1277.

(29) Fox, T.; Kollman, P. A. Application of the RESP methodology in the parametrization of organic solvents. *J. Phys. Chem. B* **1998**, *102* (41), 8070−8079.

(30) Suarez, D.; Merz, K. M. Molecular dynamics simulations of the mononuclear zinc-beta-lactamase from Bacillus cereus. *J. Am. Chem. Soc.* **2001**, *123* (16), 3759−3770.

(31) Ryde, U. Molecular-Dynamics Simulations of Alcohol-Dehydrogenase with a 4-Coordinate or 5-Coordinate Catalytic Zinc Ion. *Proteins: Struct., Funct., Genet.* **1995**, *21* (1), 40−56.

(32) Hoops, S. C.; Anderson, K. W.; Merz, K. M. Force-Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113* (22), 8262−8270.

(33) Bautista, E. J.; Seminario, J. M. Harmonic force field for glycine oligopeptides. *Int. J. Quantum Chem.* **2008**, *108* (1), 180−188.
(34) Lin, F.; Wang, R. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J. Chem. Theory Comput.* **2010**, *6* (6), 1852−1870.
(35) Peters, M. B.; Yang, Y.; Wang, B.; Fuesti-Molnar, L.; Weaver, M. N.; Merz, K. M., Jr. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6* (9), 2935−2947.
(36) Bakken, V.; Helgaker, T. The efficient optimization of molecular geometries using redundant internal coordinates. *J. Chem. Phys.* **2002**, *117* (20), 9160−9174.
(37) Case, D. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R. C. W.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, M. G. S.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, CA, 2010.
(38) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25* (2), 247−260.
(39) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.
(40) Wilson, E. B. Some Mathematical Methods for the Study of Molecular Vibrations. *J. Chem. Phys.* **1941**, *9*, 76−84.
(41) Frisch, A.; Dennington, R. D., II; Keith, T. A. *GaussView*, version 3.0; Gaussian, Inc.: Pittsburgh, PA, 2003.
(42) Frisch, M. J. T.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.1; Gaussian Inc.: Wallingford, CT, 2009.
(43) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.* **1990**, *11* (3), 361−373.
(44) Sinha, P.; Boesch, S. E.; Gu, C. M.; Wheeler, R. A.; Wilson, A. K. Harmonic vibrational frequencies: Scaling factors for HF, B3LYP, and MP2 methods in combination with correlation consistent basis sets. *J. Phys. Chem. A* **2004**, *108* (42), 9213−9217.
(45) Case, T. M. a. D. A. *Molecular Modeling of Nucleic Acids*; Leontes, N. B., SantaLucia, J., Ed.; American Chemical Society: Washington, DC, 1998; pp 379−393.
(46) Verstraelen, T. MolMod. http://molmod.ugent.be/code/wiki/MolMod (accessed Aug. 18, 2011).
(47) Ayers Group Software. http://www.chemistry.mcmaster.ca/ayers/projects.html (accessed Dec. 1, 2011),
(48) Homeyer, N.; Horn, A. H. C.; Lanig, H.; Sticht, H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.* **2006**, *12* (3), 281−289.
(49) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.* **1993**, *97* (40), 10269−10280.