JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

pubs.acs.org/JCTC

# Harmonic Force Constants for Molecular Mechanics Force Fields via Hessian Matrix Projection

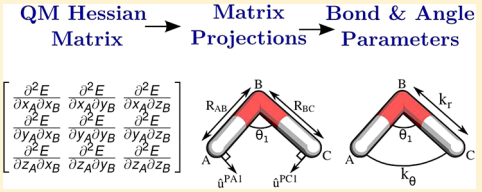
Alice E. A. Allen,[†] Michael C. Payne,[†] and Daniel J. Cole*[,‡]

[†]TCM Group, Cavendish Laboratory, 19 JJ Thomson Avenue, Cambridge CB3 0HE, United Kingdom

[‡]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom

Supporting Information

**ABSTRACT:** A modification to the Seminario method [*Int. J. Quantum Chem.* **1996**, *60*, 1271−1277] is proposed, which derives accurate harmonic bond and angle molecular mechanics force field parameters directly from the quantum mechanical Hessian matrix. The new method reduces the average error in the reproduction of quantum mechanical normal-mode frequencies of a benchmark set of 70 molecules from 12.3% using the original method, to 6.3%. The modified Seminario method is fully automated, and all parameters are computed directly from quantum mechanical data, thereby avoiding interdependency between bond and angle parameters and other components of the force field. A complete set of bond and angle force field parameters for the 20 naturally occurring amino acids is also provided for use in the future development of protein force fields.




## 1. INTRODUCTION

Molecular mechanics (MM) force fields are used to understand and predict a wide range of biological phenomena, including protein−ligand binding free energies,[1,2] enzyme catalysis,[3] and protein folding.[4] The majority of biomolecular force fields may be decomposed into intermolecular interactions, which describe the electrostatic and van der Waals energies, and intramolecular interactions, which describe covalent bonding.[5−7] The intramolecular component of the force field is typically further split into harmonic bond and angle components, which are used to describe vibrations of the bonds and angles around their equilibrium positions, and anharmonic torsional terms. Historically, in biomolecular force fields such as OPLS and AMBER, many of the bond and angle force constants were found by fitting MM normal modes and frequencies to experimental or quantum mechanical (QM) studies of small molecules.[5,8] Strictly speaking, such an approach creates interdependencies between force field parameters.[6] That is, the computed force constants are dependent on the choice of torsional and nonbonded parameters used in the original fitting procedure, and therefore changes to one component of the force field require a refit of all the other parameters. The bond force constants that could not be fit to experiment were estimated by assuming a linear relationship between the force constants and experimental bond lengths,[8] which may limit the achievable accuracy.[9] Given the importance of intramolecular interactions in determining conformational preferences of molecules[7] and reproducing accurate vibrational spectra,[10] biomolecular force field developers are beginning to reparametrize the bond and angle terms as a means to improve the accuracy of MM simulations.[7,11] However, there is no standard approach for bond and angle parametrization that combines both accuracy and ease-of-use, while removing the problem of parameter interdependence.

A number of methods have recently been developed that are aimed at finding bond and angle parameters with greater ease and accuracy.[7,9,10,12,13] These methods can be divided into fitting approaches, which rely on MM calculations as part of the parametrization process, and nonfitting approaches, which rely only on QM data. The use of multiple iterations to parametrize a MM force field through fitting to the QM Hessian matrix has been shown to give reasonably accurate MM normal modes.[10] However, the dependence of the fitting process on repeated calculations of the MM Hessian matrix results in interdependencies between force field parameters.[10] This effectively means that bond and angle parameters should be updated when changes to other components of the force field are made. Similarly, in the extensions to the CHARMM force field, which was fit to QM frequency spectra, the issue of parameter interdependencies meant that repeated parametrization of bond and angle parameters was required as dihedral and nonbonded components of the force field were updated.[6] This adds time and effort to the parametrization process.[6] Another example of a large scale parametrization approach is the method used for the AMBER ff15ipq force field.[7] Eight generations of improvements were carried out, with repeated MD simulations and QM optimization at each cycle creating tens of thousands of conformations that were used to fit the bond and angle parameters.[7] Automating this process, so that it is suitable for use by inexperienced users to parametrize molecules outside the fitting set, would not be straightforward. Speed is often a factor in fitting methods, not just because of interdependency

Received: July 20, 2017
Published: November 21, 2017

ACS Publications © XXXX American Chemical Society



DOI: 10.1021/acs.jctc.7b00785
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

but also due to difficulties in the fitting process. The Force Field Toolkit (ffTk) is a VMD plugin that works with the CHARMM force field to parametrize small molecules. This method fits the MM to the QM potential energy surface, and convergence of force constants can be slow.[13] Like all fitting approaches, this method also requires an initial estimate of the force field parameters for the first MM calculation, which relies on the preliminary values being available, and reasonably close to the optimal values. A method to fit parameters using the partial QM Hessian matrix of a molecule has recently been developed and tested on 23 molecules. This gave a mean unsigned error of 73.3 $cm^{-1}$ for the recreation of QM vibrational frequencies,[14] which gives an indication of the levels of error that are typically obtained by fitting MM parameters to QM data.

The methods discussed so far all have the disadvantages of nontransferability, interdependency of force field parameters, and reliance on an initial parameter estimate. These are inherent characteristics of methods that rely on fitting MM force field parameters to QM or experimental data. Therefore, it is seemingly advantageous to move away from using MM calculations as part of the bond and angle parametrization procedure. Nonfitting methods offer speed, transferability, and independence from the other force field components but currently lack accuracy. The most widely used of the nonfitting methods is the Seminario method,[12] which uses projections of the QM Hessian matrix to determine force constants for MM force fields and is available through the AMBER suite of programs, in the VFFDT plugin, or in the MCPB.py program. This method has been popularly applied to biomolecular systems containing metals, for which general force field parameters are typically lacking.[9,12,15−17] However, this method has been shown to be less accurate than fitting to the QM Hessian matrix for two small test sets.[10,14] In particular, the Seminario method struggled to recreate the normal modes of molecules with more than five atoms, with particular problems recreating the angle bending frequencies.[14] This points to possible inaccuracies in the angle force constants.

In this paper, we propose a modification to the Seminario method, which substantially improves the computed angle force constants by taking into account the geometry of the systems under study. We extensively test the ability of the modified Seminario method to reproduce QM vibrational frequencies and compare the accuracy to standard MM force fields and the original Seminario method. The benchmark data set comprises a total of 70 molecules, including small molecules and dipeptides, against which standard force fields have been parametrized, and also more complex organic heterocycles and a metal containing complex. For the majority of the 70 molecules tested, the modified Seminario method is more accurate than the original approach. A program that implements the method proposed is supplied, which allows users to quickly and easily derive bond and angle parameters from the output of a Gaussian09[18] frequency calculation. To prevent repetition of calculations, we have also supplied a complete set of bond and angle force field parameters for each of the 20 naturally occurring amino acids for use in future biological force fields.

## 2. THEORY

**2.1. Seminario Method.** The Seminario method was developed by Jorge Seminario in 1996[12] to parametrize harmonic bond and angle force field parameters from the QM Hessian matrix of the molecule. This provided a valuable tool for obtaining intramolecular force field parameters directly from QM data, without the need for empirical input. In this section, we outline the original Seminario methodology.

The reaction force, $\delta\mathbf{F}$, due to a small displacement $\delta\mathbf{r}$ in a system comprising $N$ atoms can be written to second order as

$$\delta\mathbf{F} = -[\mathbf{k}]\delta\mathbf{r} \tag{1}$$

where $[\mathbf{k}]$ is the $3N \times 3N$ Hessian matrix of the molecule. For practical applications in MM simulations, the relationship between the total energy of a molecule and its nuclear coordinates are typically expressed in terms of a force field equation in internal coordinates:

$$V = \sum_{\text{bonds}} \frac{1}{2}k_r(r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2}k_\theta(\theta - \theta_0)^2 + \ldots \tag{2}$$

where the first term accounts for two-body bond stretching about an equilibrium bond length ($r_0$) and the second for three-body angle bending about an equilibrium bond angle ($\theta_0$). MM force fields generally also include an anharmonic four-body torsional term, but this is not discussed further here. The objective of the Seminario method is therefore to obtain the MM harmonic force constants, $k_r$ and $k_\theta$, from the full QM Hessian matrix $[\mathbf{k}]$.

By analogy with eq 1, the force felt by atom A due to displacement of atom B is given by $\delta\mathbf{F}_A = -[\mathbf{k}_{AB}]\delta\mathbf{r}_B$. The $3 \times 3$ interatomic force constant matrix $[\mathbf{k}_{AB}]$ contains only the elements of the full Hessian matrix relating to atoms A and B:

$$[\mathbf{k}_{AB}] = -\begin{bmatrix} \frac{\partial^2 E}{\partial x_A \partial x_B} & \frac{\partial^2 E}{\partial x_A \partial y_B} & \frac{\partial^2 E}{\partial x_A \partial z_B} \\ \frac{\partial^2 E}{\partial y_A \partial x_B} & \frac{\partial^2 E}{\partial y_A \partial y_B} & \frac{\partial^2 E}{\partial y_A \partial z_B} \\ \frac{\partial^2 E}{\partial z_A \partial x_B} & \frac{\partial^2 E}{\partial z_A \partial y_B} & \frac{\partial^2 E}{\partial z_A \partial z_B} \end{bmatrix} \tag{3}$$

The three eigenvalues of $[\mathbf{k}_{AB}]$, $\lambda_i^{AB}$, are the force constants in the direction of the three eigenvectors, $\nu_i^{AB}$. However, we instead require the force constants for changes in intramolecular bond lengths and angles. To calculate the bond force constant for the bond AB, each eigenvector is projected onto the direction of the bond vector, $\hat{u}^{AB}$:

$$k_r = \sum_{i=1}^{3} \lambda_i^{AB} |\hat{u}^{AB} \cdot \hat{\nu}_i^{AB}| \tag{4}$$

In the original Seminario paper the definition of a bonded atom was determined by the eigenvalues of $[\mathbf{k}_{AB}]$. We have not used this definition and use the conventional definition of bonded atoms specified by the force field.

The angle force constant, $k_\theta$, is more complex as it involves projections onto directions perpendicular to two different bonds AB and CB. Let us define two vectors, $\hat{u}^{PA1}$ and $\hat{u}^{PC1}$ (Figure 1a), that are perpendicular to the bonds AB and CB, respectively, and lie in the plane ABC. Then $k_{PA}$ and $k_{PC}$ are defined as the corresponding force constants obtained by projecting the eigenvectors of the partial Hessian matrix onto these two vectors:


DOI: 10.1021/acs.jctc.7b00785
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

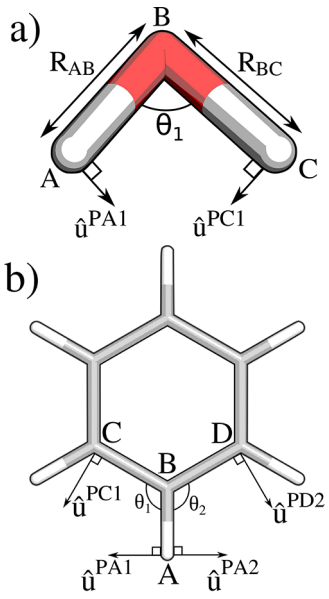



**Figure 1.** (a) Original Seminario method applied to a water molecule. (b) Extension of the original Seminario method to a larger molecule (benzene). If atom A moves in the direction perpendicular to bond AB, $\hat{u}^{PA1}$, both angles $\theta_1$ and $\theta_2$ are altered. Therefore, the angle force constant obtained via projection of the eigenvector of the matrix $[\mathbf{k}_{AB}]$ onto $\hat{u}^{PA1}$ is overestimated by a factor of 2.

$$k_{PA} = \sum_{i=1}^{3} \lambda_i^{AB} |\hat{u}^{PA1} \cdot \hat{v}_i^{AB}| \tag{5}$$

$$k_{PC} = \sum_{i=1}^{3} \lambda_i^{CB} |\hat{u}^{PC1} \cdot \hat{v}_i^{CB}| \tag{6}$$

Via analogy to two springs connected in series, the angle force constant is then approximated by

$$\frac{1}{k_\theta} = \frac{1}{R_{AB}^2 k_{PA}} + \frac{1}{R_{CB}^2 k_{PC}} \tag{7}$$

where $R_{AB}$ and $R_{CB}$ are the two bond lengths (Figure 1a).

**2.2. Modified Seminario Method.** The Seminario method for the derivation of harmonic angular force constants assumes that the change in energy associated with the displacement of atom A along the direction $\hat{u}^{PA1}$ will only change the angle involving atoms A, B, and C as in Figure 1a. However, in larger molecules, neighboring angles may also be altered by a displacement of atom A in the direction of $\hat{u}^{PA1}$. Let us consider how the Seminario method would calculate the angle force constants involving four atoms (A, B, C, D) in the same plane, such as in a benzene molecule (Figure 1b). The Seminario method finds the force constant for angle ABC, $\theta_1$, from the projections of the eigenvectors of the partial Hessian matrices on to $\hat{u}^{PA1}$ and $\hat{u}^{PC1}$. Hence the Seminario estimate of $k_{\theta_1}$ includes all QM forces on atom A acting in the direction $\hat{u}^{PA1}$. Importantly, however, the equivalent calculation of $k_{\theta_2}$ also includes all QM forces on atom A acting in the direction $\hat{u}^{PA1}$. This inevitably results an intramolecular MM force field that is too stiff.

We can think of this problem in terms of the change in energy caused by a small change, $\Delta x$ along $\hat{u}^{PA1}$, that would be computed using the original Seminario method:

$$k_{PA1}^{Seminario}(\Delta x)^2 = k_{PA1}(\Delta x)^2 + k_{PA2}(\Delta x)^2 \tag{8}$$

where $k_{PA1}$ ($k_{PA2}$) is the hypothetical value of $k_{PA1}^{Seminario}$ ($k_{PA2}^{Seminario}$) that would be computed if ABC (ABD) existed in isolation from all other angles. For the water example $k_{PA1}^{Seminario} = k_{PA1}$. For the benzene molecule, $k_{PA1} = k_{PA2}$ and so the Seminario method overestimates the change in energy by a factor of 2.

If ABC and ABD are not in the same plane, movement in the direction $\hat{u}^{PA1}$ can still cause displacement in the direction of neighboring angles. The change in energy predicted by the original Seminario method may then be approximated by

$$k_{PA1}^{Seminario}(\Delta x)^2 = k_{PA1}(\Delta x)^2 + k_{PA2}(\Delta x |\hat{u}_{PA1} \cdot \hat{u}_{PA2}|)^2 \tag{9}$$

Assuming further that $k_{PA1} \approx k_{PA2}$, which is true when both angles are in the same plane:

$$k_{PA1}^{Seminario} = k_{PA1}(1 + |\hat{u}_{PA1} \cdot \hat{u}_{PA2}|^2) \tag{10}$$

$$k_{PA1} = \frac{k_{PA1}^{Seminario}}{(1 + |\hat{u}_{PA1} \cdot \hat{u}_{PA2}|^2)} \tag{11}$$

Equation 11 rescales the original value of $k_{PA1}^{Seminario}$ by a factor that accounts for the geometry of the molecule. To extend the above analysis to sites B with multiple angles, we have found empirically that the mean of the additional contribution ($|\hat{u}_{PA1} \cdot \hat{u}_{PA2}|^2$) from all neighboring angles gives the most reasonable agreement with the QM vibrational frequency spectra. This results in our modified formula for the angle force constant for ABC when $N > 1$ and $M > 1$, where $N$ ($M$) is the number of angles in the force field that have a central atom B and involve movement of the bond AB (BC):

$$\frac{1}{k_\theta} = \frac{1 + \frac{\sum_{i=1}^{N} |\hat{u}^{PA_1} \cdot \hat{u}^{PA_i}|^2 - 1}{N-1}}{R_{AB}^2 k_{PA}^{Seminario}} + \frac{1 + \frac{\sum_{i=1}^{M} |\hat{u}^{PC_1} \cdot \hat{u}^{PC_i}|^2 - 1}{M-1}}{R_{CB}^2 k_{PC}^{Seminario}} \tag{12}$$

If $N = 1$ ($M = 1$) the left (right) hand component is replaced by the original Seminario method.

**2.3. Computational Implementation.** Our proposed modification to the Seminario method comprises eq 4 for the intramolecular harmonic bond force constants and eq 12 for the modified angle force constants. The equilibrium bond lengths and angles are obtained from the optimized QM structure. The modified Seminario method is available for use through a MATLAB program, which may be freely downloaded from https://github.com/aa840/ModSeminario along with a short tutorial explaining how to use the program to find the bonded parameters of a benzene molecule. For the example given, the Hessian matrix of the molecule can be converted into bond and angle parameters in a matter of seconds on a standard desktop computer. Larger molecules may also be parametrized in negligible computed times because the method scales approximately linearly with the number of bonds and angles.

The optimized structure and connectivity of the molecule, as well as the QM Hessian matrix, is read in from Gaussian 09[18] output files (specifically .fchk and .log files). Optionally, a BOSS **z**-matrix,[19] which can be produced using the LigParGen web server,[20−22] may be supplied to provide the OPLS atom types. If a **z**-matrix is supplied as input, the OPLS atom types are used to return the average value for each bond and angle class. However, if OPLS atom types are not required, or are unavailable (for example, for molecules containing a metal), the Gaussian 09 output files can be used in isolation, with no bond


DOI: 10.1021/acs.jctc.7b00785
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

and angle parameter averaging performed. Thus, the program can be used for a wide range of molecules and force fields.

Following standard practice, the QM-derived vibrational frequencies of a molecule can be multiplied by a constant to better fit experimental vibrational spectra.[23] This is incorporated into the modified Seminario method by multiplying the bond and angle force constants by the square of the frequency scaling constant.[24] This scaling constant can be altered by the user according to the level of QM theory employed, or set equal to one.

## 3. SIMULATION METHODS

To test the accuracy of the modified Seminario method, 38 small organic molecules were chosen with a diverse range of chemical structures. Test sets of this nature are commonly used to parametrize MM biomolecular force fields. The molecules contained more than six atoms to ensure that the effect of our angle correction is apparent. Following the small molecule validation set, we also repeated our analysis on a set of ten hetereocyclic molecules. The full list of small and heterocyclic molecules is provided in section S1 in the Supporting Information.

For each molecule, a structural optimization and frequency calculation was performed using Gaussian 09 with the $\omega$B97XD functional and a 6-311++G(d,p) basis set.[18] The QM vibrational frequencies were rescaled by a factor of 0.957, which is the value recommended for the $\omega$B97XD/6-311G-(d,p) level of theory by the Computational Chemistry Comparison and Benchmark DataBase.[25] The same scaling factor is also used to effectively scale the MM frequencies, as outlined in section 2.3. The level of QM theory chosen for the frequency calculation is the same as that used in the recent reparametrization of the protein backbone torsional parameters for OPLS-AA/M.[26] To ensure that this choice did not significantly influence results, our analysis was repeated for a subset of 10 small molecules using the B3LYP/cc-pVTZ level of theory. As reported in Table S4, the computed accuracy of our method is not strongly dependent on the choice of underlying QM data.

Bond and angle parameters for each molecule were found, as described above, from the computed QM Hessian matrices, using the original and modified Seminario method. Dihedral and nonbonded parameters were assigned from the OPLS/CM1A force field[21,27] using the LigParGen web server to obtain **z**-matrices.[20] The MM normal modes and frequencies were calculated using the BOSS general purpose molecular modeling software with Broyden–Fletcher–Goldfarb–Shanno (BFGS) structural optimization.[19] The mean percentage error in each molecule is computed as $\frac{100}{3N-6}\sum_{i=7}^{3N}\left|\frac{\alpha\nu_i^{QM}-\nu_i^{MM}}{\alpha\nu_i^{QM}}\right|$ where $\nu_i^{MM/QM}$ is the frequency of the $i$th MM/QM normal mode, $\alpha$ is the vibrational scaling factor, and $N$ is the number of atoms in the molecule. The mean unsigned error (MUE) is also given for comparison and is computed as $\frac{1}{3N-6}\sum_{i=7}^{3N}|\alpha\nu_i^{QM}-\nu_i^{MM}|$. Although the above measures of error are commonly used in assessing the accuracy of bond and angle parameters, we emphasize that the QM frequencies, which are treated as "ideal" values, are derived from the same QM Hessian matrix used to parametrize the bond and angle terms.

To provide direct comparison with the MCPB.py force field parametrization program,[9] we also analyzed the complex tris(1,10-phenanthroline)-osmium(II) ($Os[(phen)_3]^{2+}$) shown in Figure 2.[28] The QM Hessian matrix, computed using the B3LYP functional and a 6-31G(d) basis set, was obtained

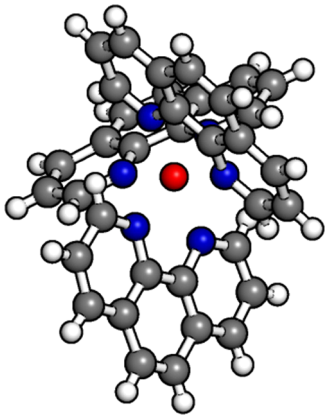


**Figure 2.** $Os[(phen)_3]^{2+}$ complex. Hydrogen atoms are white, carbon gray, nitrogen blue, and osmium red.

directly from the work by Li et al.[9] Bond and angle parameters were computed using the modified Seminario method. For consistency with the MCPB.py analysis, we computed MM normal modes using AMBER16,[29] with the AMBER ff14SB and GAFF force fields for torsional and nonbonded parameters. A vibrational scaling factor was not applied in this case.

Finally, we computed bond and angle force field parameters for each of the 20 naturally occurring amino acids with the same methods used for the small molecule and heterocycle data sets. The OPLS-AA/M force field was used for all torsional and nonbonded parameters.[26] The amino acids (X) were blocked with acetyl and *N*-methyl groups (Ace–X–NMe). We use this structure as our definition of a dipeptide (a single amino acid with two peptide bonds). A total of 80 structures were analyzed to account for variation in backbone and side chain conformations. A minimum of one $\beta$-sheet and one $\alpha$-helical conformation was tested for each dipeptide, which have $\psi$ and $\phi$ dihedral angles of (−60°, −45°) and (−135°, +135°) respectively.[26] Additional starting configurations for larger amino acids were generated by fixing the backbone dihedral angles and scanning the side chain dihedral angle (N–C$\alpha$–C$\beta$–X$\gamma$, where the atom type X$\gamma$ depends on the amino acid) for local minima. The starting structures were fully optimized as part of the QM frequency calculation, ensuring a representative sampling of low energy structures. The reported bond and angle force field parameters were averaged over the different conformations produced.

## 4. RESULTS

Table 1 summarizes the ability of various force field parametrization techniques to reproduce the QM vibrational frequencies of a range of tested molecules. The full list of molecules and their associated errors are given in section S1, along with examples of the vibrational spectra for four molecules. Focusing first on the small molecule data set, the percentage error for the vibrational frequencies computed using the OPLS force field is 7.3%. As expected, the computed frequencies are very similar to high level QM data because force fields are parametrized to fit experimental vibrational spectra of small molecules such as these.[5,8] Similar accuracy is expected for other standard MM biomolecular force fields, which often employ the same bond and angle parameters. Next, we



DOI: 10.1021/acs.jctc.7b00785
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

**Table 1. Mean Percentage Error (%) in the MM Vibrational Frequencies for OPLS and the Original/Modified Seminario Parametrization Schemes**[a]

| | OPLS | original Seminario | modified Seminario |
|---|---|---|---|
| small molecules | 7.3% (60.4) | 12.3% (119.5) | 6.4% (52.3) |
| heterocycles | 8.6% (82.6) | 11.7% (132.3) | 6.8% (52.8) |
| dipeptides | 7.0% (46.6) | 12.4% (104.3) | 6.1% (39.5) |
| average | 7.4% (59.4) | 12.3% (116.7) | 6.3% (48.5) |

[a]The value shown in brackets is the mean unsigned error ($cm^{-1}$). The QM frequencies used in the calculation of the error have been scaled to better reproduce experimental frequencies.[23]

reparameterize the bond and angle equilibrium values and force constants using the method proposed by Seminario[12] and combine the parameters with OPLS torsional and nonbonded parameters. The resulting error is almost twice as high as that of the standard force field. Mean unsigned errors of >100 $cm^{-1}$ have been reported in other studies,[10,14] which casts doubt on the suitability of the Seminario method as an automated parametrization tool.[9] In contrast, the error in our modified parametrization scheme (6.4%) is much lower than the original Seminario method and similar to the OPLS force field. The corrections that we have made to the Seminario method, described in section 2.2, result in a more accurate recreation of the QM vibrational spectra.

As discussed, standard MM force fields are expected to perform well for this small molecule data set. As a more stringent test, we have computed the QM vibrational spectra of ten more complex heterocyclic compounds, which are expected to be less structurally similar to the original parametrization set. Table 1 summarizes the average error in the three parametrization methods across all ten molecules, and Figure 3 further breaks down the results by molecule. As expected, the error in the vibrational frequencies computed using OPLS (8.6%) is higher than that computed for the small molecule test set. However, the original Seminario yields even higher errors, again indicating its unsuitability for force field parametrization. Encouragingly, the modified Seminario method maintains a low error (6.8%), which is largely constant across the heterocycle test set and is consistently lower than the original Seminario method (Figure 3). Closer examination reveals that the majority of the improvement in the accuracy of the normal modes is brought about by the changes to the harmonic force constants, rather than the bond lengths or equilibrium angles. Some parameters have very large deviations from the corresponding OPLS parameters. For example, one of the C−C bond force constants in pyrrole, found using the modified Seminario method, is 25% lower than the corresponding OPLS parameter. Improvements in the optimized structures are also observed with the new parameters for heterocyclic molecules, particularly the four-membered rings. In the QM optimized structures, all the heavy atoms in $\beta$-lactam and oxetane are coplanar, which is correctly reproduced by the modified Seminario parameters. However, optimization with OPLS yields slightly twisted structures, with a computed C−O−C−C dihedral angle of 20.2° for oxetane and a C−N−C−C dihedral angle of 11.5° in $\beta$-lactam.

Figure 4 shows a comparison between the QM and MM vibrational frequency spectra of $Os[(phen)_3]^{2+}$. The MCPB.py

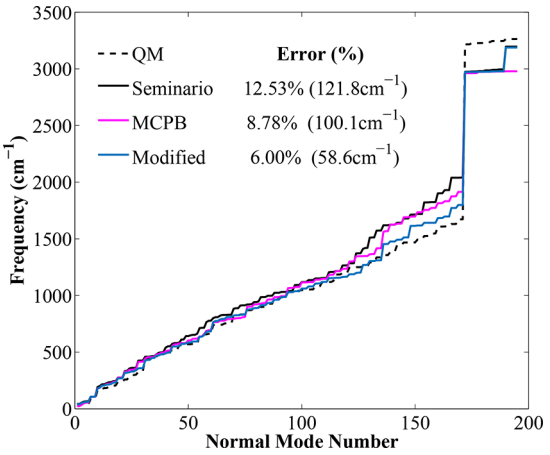



**Figure 4.** Vibrational spectrum of $Os[(phen)_3]^{2+}$ computed using QM and compared with the original and modified Seminario methods, as well as the bonded parameters reported in ref 9 (MCPB). The mean percentage error for each method is given in the key. No vibrational scaling factor has been applied to the QM frequencies.

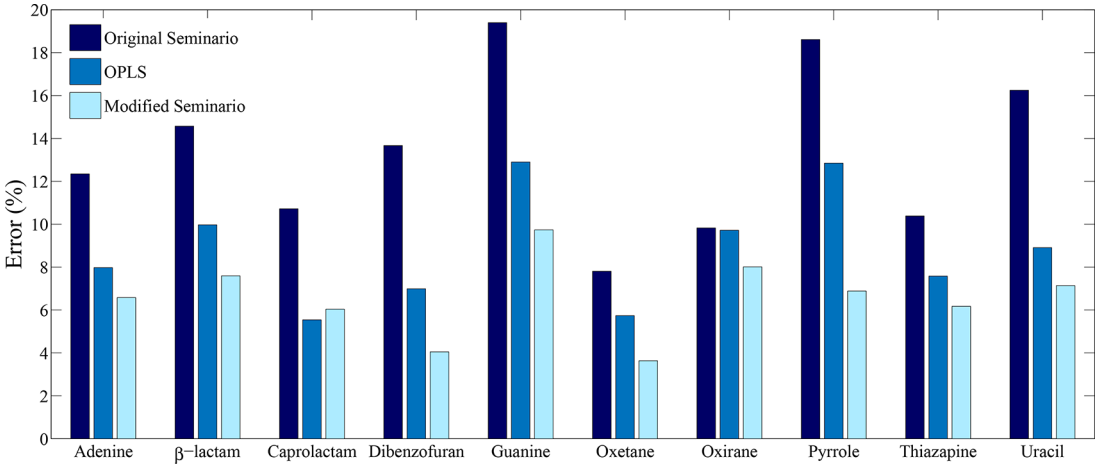



**Figure 3.** Error in the MM vibrational frequencies for a set of heterocyclic molecules, using bond and angle parameters from the original Seminario method, OPLS, and the modified Seminario method. The QM frequencies used in the calculation of the error have been scaled to better reproduce experimental frequencies.[23]



DOI: 10.1021/acs.jctc.7b00785
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

method uses the original Seminario method to parametrize bonds and angles involving the metal ion and applies the standard AMBER force field elsewhere.[9] For direct comparison, we have used identical dihedral and nonbonded parameters but replaced all bond and angle parameters with those computed using the original and modified Seminario methods. All methods agree well in the low frequency range 0−1250 $cm^{-1}$, whereas reproduction of the very high frequency (>2500 $cm^{-1}$) vibrations of bonds involving hydrogen is problematic for all methods. The original Seminario method and MCPB methods clearly overestimate the vibrational frequencies of normal modes in the intermediate regime (1250−2500 $cm^{-1}$). These modes largely involve angle bending motions, which are precisely the motions that we set out to correct in the modified Seminario method. The overall error (6.0%) of our modified Seminario method is half that computed using the original Seminario method, and very similar to the errors computed for the small molecule and heterocycle data sets. Furthermore, this additional test case demonstrates that the modified Seminario method works well for relatively large system sizes (67 atoms) and is not too dependent on the underlying force field that is used to compute torsional and nonbonded energetics.

The validation tests described in this paper reveal that harmonic force constants derived using the modified Seminario method give vibrational spectra that are in very good agreement with QM data across a wide range of molecules. We therefore envisage this method being used as a toolkit for automated parametrization of molecules that are missing force field parameters (for example, metal complexes), or for which the transferability of the standard force field parameters are questionable (for example, heterocyclic molecules). As a further resource, and also to test whether the modified Seminario method is suitable for *transferable* force field parametrization, we have computed a new bond and angle parameter set for the 20 naturally occurring amino acids using our new method. The amino acids are blocked with acetyl and *N*-methyl groups to form dipeptides, as described in section 3. We performed QM calculations on a total of 80 dipeptide structures (including different backbone and side chain conformations) and averaged parameters for each atom type over all structures and amino acids. The resulting parameter set is given in the Supporting Information. Using these averaged bond and angle parameters alongside the OPLS-AA/M force field, we computed the vibrational spectra of each of the 20 amino acids and compared our results with QM data. The results are summarized in Table 1, and show the expected trend. The OPLS force field, which has been parametrized to reproduce experimental vibrational spectra of small molecules, has a low percentage error of 7.0% relative to our QM data. The original Seminario method fails to reproduce the QM vibrational spectra, whereas the modified method results in a slightly lower error (6.1%) than the OPLS force field. As a comparison, we also computed the vibrational spectra for each amino acid using bond and angle parameters that are *specific* to that molecule (Table S3). However, the error is virtually identical to the averaged parameter set for the modified Seminario method, indicating that the harmonic bond and angle parameters are indeed transferable.

To better understand how the bond and angle parameters vary, the modified Seminario parameters were compared to the OPLS force field (section S2). It was found that the bond lengths and equilibrium angles do not deviate far from the OPLS parameters. However, the modified Seminario bond force constants are generally lower than the OPLS parameters, with the mean bond force constant being 411 (kcal/mol)/$Å^2$ for OPLS and 341 (kcal/mol)/$Å^2$ for our new parameter set. In contrast, the modified Seminario angle force constants are slightly higher than for the OPLS force field. However, even more apparent is the larger range of force constants that are computed using the modified Seminario method. OPLS angle force constants range between 33−85 (kcal/mol)/$rad^2$ for the amino acid set, whereas the corresponding modified Seminario parameters lie between 22−178 (kcal/mol)/$rad^2$ (Figure S12).

## 5. DISCUSSION AND CONCLUSION

We have developed a method for the parametrization of harmonic bond and angle force constants for molecular mechanics force fields. The method recreates QM normal-mode frequencies with a consistently high level of accuracy, and uses no empirical or MM data in the parametrization process. Use of our bonded parameters results in similar levels of error to standard force fields for a general set of small molecules and dipeptides, and a noticeable improvement for heterocyclic molecules. In certain cases, the optimized structures of hetereocyclic molecules are greatly improved using the new approach. The parameters have been computed using a modified version of the widely used Seminario method,[12] in which critical improvements have been made to the angle force constants. For the majority of the 70 molecules tested, the modified Seminario method is more accurate than the original approach.

Although the accuracy of the modified Seminario method is extremely good, Figure 4 reveals possible areas for further improvement. All methods tested show quite large errors in the very high frequency bond stretching modes involving hydrogen (although these modes are unlikely to critically affect many computed properties of interest). The modified Seminario method substantially improves the recreation of intermediate modes involving angle-bending motions. It should be emphasized that we do not claim that eq 12 is the only method for partitioning $k_\theta$ parameters from the full QM Hessian matrix, and other schemes are possible. In fact, we investigated one such scheme during development of the modified Seminario method. Motivated by the observation that the original Seminario method strongly overestimates the stiffness of larger molecules, we investigated a simple rescaling of the angle force constants by a constant multiplicative factor (Section S3). This method gave percentage errors of around 8.5%, which is an improvement over the original Seminario method, but significantly worse than our modified approach, which accounts more rigorously for the molecular environment.

With regard to the low frequency portion of the vibrational spectra, in this paper, we have combined the derived harmonic bond and angle parameters with torsional and nonbonded parameters from standard MM force fields. Figure 4 is typical of the vibrational spectra computed in this study and shows that the errors in the low frequency part of the spectrum are low. Nevertheless, further improvements in accuracy could be possible by reparameterizing the torsional terms using the modified Seminario bond and angle parameters. Finally, we have assumed throughout this study that the QM normal modes and frequencies are an accurate representations of experimental values. This is reliant on the vibrational scaling factors used being suitable for the molecules tested, and not being frequency dependent.[30]

The Seminario method is one of a number of methods that can be used to parametrize harmonic bond and angle force field


DOI: 10.1021/acs.jctc.7b00785
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

terms.[7,10,14] The level of accuracy that can be obtained by fitting MM parameters to the QM Hessian matrix has been previously reported as 63.9 or 73.3 $cm^{-1}$ depending on the details of the fitting procedure.[10,14] These methods appear to be less accurate than the modified Seminario method, though they have only been tested on a small number of molecules, and further testing should ideally be carried out on equivalent data sets with identical error analysis. As well as potentially improved accuracy, the modified Seminario method also has other clear advantages over all fitting methods currently in use. Because the force field parameters are derived directly from the QM Hessian matrix, initial estimates of the remaining force field parameters are not required, and interdependencies between the different components of the force field are avoided. Reduction of parameter interdependencies is desirable to prevent the need for several iterations of fitting, and therefore to produce the most efficient parametrization schemes. Recent efforts to improve biomolecular force fields have seen a number of groups reparametrize the bond and angle components of proteins using fitting approaches.[7,11] Therefore, each new iteration of these force fields will require a full reparametrization of the bonded terms. We offer a simple, alternative solution by supplying a library of bond and angle parameters for the set of 20 naturally occurring amino acids. These parameters can then be used as the basis for any future protein force fields that employ the standard harmonic functional form for bond and angle terms.

The modified Seminario method has been implemented in a freely available program and offers a means to parametrize bond and angle terms in a fast and automated way. Future work will aim to combine the modified Seminario method with automated fitting of torsional parameters. Developments such as these will be crucial in our overall goal of creating an automated workflow for the accurate parametrization of biomolecular MM force fields directly from QM data.[31−33]

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00785.

List of all molecules in each test set and their associated errors and vibrational spectra, comparisons between the modified Seminario parameters and OPLS, and a discussion of the scaled Seminario method including mean percentage errors between the QM and MM normal mode frequencies and a plot of modified vs original Seminario parameters (PDF)

Bond and angle force field parameters for 20 amino acids computed using the modified Seminario method (ZIP)

## ■ AUTHOR INFORMATION


### Corresponding Author

*D. J. Cole. E-mail: daniel.cole@ncl.ac.uk.

### ORCID

Alice E. A. Allen: 0000-0002-8727-8333

Daniel J. Cole: 0000-0003-2933-0719


### Notes

The authors declare no competing financial interest.

The program implementing the modified Seminario method can be freely downloaded at https://github.com/aa840/ModSeminario with a tutorial demonstrating benzene parametrization.

## ■ ACKNOWLEDGMENTS

The authors are grateful to Joshua Horton (Newcastle University) for helpful discussions and Michael Robertson (Yale University) for providing amino acid input files. A.E.A.A. acknowledges financial support from the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science under grant EP/L015552/1.

## ■ REFERENCES


(1) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695−2703.

(2) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. Molecular dynamics and Monte Carlo simulations for protein−ligand binding and inhibitor design. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 966−971.

(3) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational enzymology. *Chem. Commun.* **2010**, *46*, 2354−2372.

(4) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.

(5) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(6) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(7) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12*, 3926−3947.

(8) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(9) Li, P.; Merz, K. M. MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599−604.

(10) Burger, S. K.; Lacasse, M.; Verstraelen, T.; Drewry, J.; Gunning, P.; Ayers, P. W. Automated Parametrization of AMBER Force Field Terms from Vibrational Analysis with a Focus on Functionalizing Dinuclear Zinc(II) Scaffolds. *J. Chem. Theory Comput.* **2012**, *8*, 554−562.

(11) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023−4039.

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

(13) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* **2013**, *34*, 2757−2770.

(14) Wang, R.; Ozhgibesov, M.; Hirao, H. Partial Hessian fitting for determining force constant parameters in molecular mechanics. *J. Comput. Chem.* **2016**, *37*, 2349−2359.

(15) Lin, F.; Wang, R. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J. Chem. Theory Comput.* **2010**, *6*, 1852−1870.

(16) Bautista, E. J.; Seminario, J. M. Harmonic force field for glycine oligopeptides. *Int. J. Quantum Chem.* **2008**, *108*, 180−188.
(17) Zheng, S.; Tang, Q.; He, J.; Du, S.; Xu, S.; Wang, C.; Xu, Y.; Lin, F. VFFDT: A New Software for Preparing AMBER Force Field Parameters for Metal-Containing Molecular Systems. *J. Chem. Inf. Model.* **2016**, *56*, 811−818.
(18) Frisch, M. J.; Trucks, G. W.; 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, J. M.; 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 E.01; Gaussian Inc.: Wallingford, CT, 2009.
(19) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689−1700.
(20) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331−W336.
(21) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864−3870.
(22) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6665−6670.
(23) Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502−16513.
(24) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. An automatic method to generate force-field parameters for hetero-compounds. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 274−289.
(25) Johnson, R. D., III. *NIST Computational Chemistry Comparison and Benchmark Database*, accessed 2016; http://cccbdb.nist.gov/.
(26) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.
(27) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J. Phys. Chem. B* **2004**, *108*, 16264−16270.
(28) Demadis, K.; Dattelbaum, D.; Kober, E.; Concepcion, J.; Paul, J.; Meyer, T.; White, P. Vibrational and structural mapping of $[Os(bpy)^3]^{3+/2+}$ and $[Os(phen)^3]^{3+/2+}$. *Inorg. Chim. Acta* **2007**, *360*, 1143−1153.
(29) Case, D.; Cerutti, D.; Cheatham, T. D., III; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Simmerling, C.; Botello-Smith, W.; Swails, J.; Walker, R.; Wang, J.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. *AMBER 2017*; University of California: San Francisco, 2017.
(30) Irikura, K. K.; Johnson, R. D.; Kacker, R. N. Uncertainties in Scaling Factors for ab Initio Vibrational Frequencies. *J. Phys. Chem. A* **2005**, *109*, 8430−8437.
(31) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312−2323.
(32) Lee, L. P.; Limas, N. G.; Cole, D. J.; Payne, M. C.; Skylaris, C.-K.; Manz, T. A. Expanding the Scope of Density Derived Electrostatic and Chemical Charge Partitioning to Thousands of Atoms. *J. Chem. Theory Comput.* **2014**, *10*, 5377−5390.
(33) Lee, L. P.; Cole, D. J.; Skylaris, C.-K.; Jorgensen, W. L.; Payne, M. C. Polarized Protein-Specific Charges from Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2013**, *9*, 2981−2991.


DOI: 10.1021/acs.jctc.7b00785
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX