Quantum hessian fitting for quick determination of force constant parameters in molecular mechanics

Wang, Ruixing

2017

Wang, R. (2017). Quantum hessian fitting for quick determination of force constant parameters in molecular mechanics. Master's thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/70252>

<https://doi.org/10.32657/10356/70252>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**Quantum Hessian Fitting for Quick Determination of
Force Constant Parameters in Molecular Mechanics**

WANG RUIXING

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2017

Quantum Hessian Fitting for Quick Determination of Force Constant Parameters in Molecular Mechanics

WANG RUIXING

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A thesis submitted to the
Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Science

2017

Abstract

We present three new methods called Full, Partial, and Internal Hessian Fitting (FHF, PHF, and IHF) for deriving force constant parameters that are used in molecular mechanics (MM) force fields to describe the bond-stretching, angle-bending, dihedral-torsion, and improper-torsion terms. The MM-calculated Hessian matrices are made as close as possible to the QM-calculated ones. The Hessian fitting processes are done analytically and thus rapidly, yielding force constant parameters as the output. We herein apply our methods to derive force constant parameters for the AMBER-type energy expression. Test calculations on several different molecules show good performance of the parameter sets produced by our methods in terms of how well they can reproduce QM-calculated frequencies. We also notice that the nonbonded interactions sometimes overwhelm the bonded ones, resulting in distorted geometries. This problem is significant when soft bonds are involved in the target molecule as in the case of secondary building units of metal-organic frameworks, where the MM-optimized geometry sometimes deviates significantly from the QM-optimized one. We show that this problem is rectified effectively by use of a simple procedure called Katachi that modifies the equilibrium bond distances and angles in bond-stretching and angle-bending term.

Acknowledgements

I would first like to express my sincere gratitude to my supervisor Asst. Prof. Hajime Hirao for his encouragement and supports. He is a patient, supportive and impressive supervisor who gave me many precious suggestions and instructions. The whole research cannot even be possible without him.

I also would like to thank Assoc. Prof. Mu Yuguang and Dr. Lu Yunpeng for their constructive comments and suggestions to my research, which encouraged me to broaden and deepen the scope of my research and investigation in depth. In addition, I would like to thank Prof. Lee Soo Ying for his lectures that helped me to understand the quantum chemistry and this knowledge helped me a lot in my research.

I appreciate the cheerful discussions and suggestions from my labmates, and all the precious time I shared with them. I also thank my roommates, with whom I spent a very pleasant and memorable time.

Finally, I must express my very profound gratitude to my parents and to my beloved girlfriend Sherry for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. Thank you.

Table of Contents

Abstract.....	i
Acknowledgements	ii
Table of Contents.....	iii
List of Figures	iv
List of Tables.....	v
List of Abbreviations	vi
1. Introduction.....	1
1.1 Molecular Mechanics, Force Field and Parameterization	1
1.2 Previous Works	2
2. Methods.....	3
2.1 Hessian Fitting.....	3
2.1.1 Full Hessian Fitting (FHF)	5
2.1.2 Partial Hessian Fitting (PHF)	6
2.1.3 Internal Hessian Fitting (IHF)	9
2.2 Amendment of Equilibrium Values (Katachi)	11
3. Implementation	11
3.1 Sample molecules, QM and MM Calculations	11
3.2 Evaluating Performance	13
4. Results and Discussion	14
4.1 Test of Methods	14
4.2 Test of Katachi Amendment	16
4.3 Discussion	18
5. Conclusion	20
Appendices	21
References	23
List of Publications and Presentations	26

List of Figures

- Figure 2.1 Schematic illustration of a chain segment within a molecule
- Figure 2.2 a) Two dihedral angles including atoms i and l in a six-membered ring.
b) A dihedral angle and an angle in a five-membered ring.
c) An improper angle and an angle
- Figure 3.1 Sample molecules considered in this study
- Figure 4.1 Superposition of the QM- (red) and MM- (yellow) optimized geometries of a) **21**, b) **22**, and c) **23**. Here, the force constant parameters for MM (with $s_{ij}^{es} = 0.0$ for 1–4 interactions) were determined by the PHF method.
- Figure 4.2 Superposition of the QM (red) and MM (yellow) optimized geometries of a) **21**, b) **22**, and c) **23**. Here, the force constant parameters and the equilibrium bond distances and angles for MM (with $s_{ij}^{es} = 0.0$ for 1–4 interactions) were determined by the PHF-K method.
- Figure 4.3 Comparison of the values for molecules **1–23**. MM data were calculated using different methods; for molecules **5**, **9**, **10**, **14–18**, and **21–23**, Katachi amendment was applied.

List of Tables

Table 4.1 Frequencies (cm^{-1}) for hydrogen peroxide calculated using different methods and their deviations from QM frequencies (Δ_{freq} and $\frac{\Delta_{freq}}{N_q}$, in cm^{-1}).

Table A1 Comparison of Δ_{freq} (cm^{-1}), $\frac{\Delta_{freq}}{N_q}$ (cm^{-1}), and rmsd values ($\text{\AA}$) for Molecules 1-20 Obtained with Different Methods.

Table 4.2 Comparison of Δ_{freq} (cm^{-1}), $\frac{\Delta_{freq}}{N_q}$ (cm^{-1}), and rmsd values ($\text{\AA}$) for MOF-SBUs 21–23 Obtained with the PHF and Seminario Methods.

Table A2 Comparison of Δ_{freq} (cm^{-1}), $\frac{\Delta_{freq}}{N_q}$ (cm^{-1}), and rmsd values ($\text{\AA}$) Obtained with Different Methods and Katachi Amendment

Table 4.3 Comparison of Δ_{freq} (cm^{-1}), $\frac{\Delta_{freq}}{N_q}$ (cm^{-1}), and rmsd values ($\text{\AA}$) for MOF-SBUs 21–23 Obtained with the PHF and Seminario Methods with Katachi Amendment

List of Abbreviations

QM	Quantum mechanics or quantum mechanical
MM	Molecular mechanics or molecular mechanical
MOF	Metal organic framework
SBU	Secondary building blocks
vdW	van der Waals
GA	Genetic Algorithm
GAFF	General AMBER force field
FHF	Full Hessian Fitting
PHF	Partial Hessian Fitting
IHF	Internal Hessian Fitting

1. Introduction

1.1 Molecular Mechanics, Force Field and Parameterization

The high computational demand of quantum mechanics (QM) based calculations prohibits the QM calculation for large biomolecular systems, such as proteins, nucleic acids, cell membranes, and other systems such as liquids and metal-organic frameworks (MOFs). Therefore, computationally much less demanding molecular mechanics (MM) based simulations have been widely used for these systems. MM provides potential energy surfaces of molecules in empirical, effective and efficient manners and is often used with geometry optimization, molecular dynamics, and Monte Carlo methods.¹⁻¹⁰ A certain set of MM functions and associated parameters is also called a force field. Various MM force fields have been developed so far. For example, in biomolecular simulations, multiple versions of AMBER force field have been developed, which commonly has the following form of potential energy function¹¹⁻¹⁵:

$$V^{MM} = \sum_{bonds} k_b (r - r_{eq})^2 + \sum_{angles} k_a (\theta - \theta_{eq})^2 + \sum_{dihedrals} \sum_n k_{d_n} [1 + \cos(n\phi - \gamma_n)] + \sum_{impropers} k_{im} [1 + \cos(n\phi - \gamma)] + \Phi^{nb} \quad (1.1)$$

where the first to fifth terms in Eq.(1.1) account for the bond-stretching, angle-bending, dihedral-torsion, improper-torsion, and nonbonded contributions, respectively. Parameters k_b and k_a are the force constants used for describing the bond-stretching and angle-bending interactions by quadratic functions, respectively, while r_{eq} and θ_{eq} are their respective equilibrium values. Dihedral torsions are described by Fourier series. In many simple cases, one Fourier term is used in AMBER to describe a dihedral term, and we now consider such cases. Parameter n is the periodicity and γ is the phase of a Fourier term.

Factor k_{d_n} is usually denoted $\frac{V_n}{2}$, where V_n is usually called the barrier height; but for convenience, we use k_{d_n} here and call it force constant of a dihedral angle. The improper torsion terms are not always used but could be important in certain cases where the torsion should be described but a dihedral angle cannot be defined from bonded atoms. The definition of an improper term is identical to that of a dihedral term, but the former term has only one Fourier term for $n = 2$. The last term in Eq.(1.1) denotes the pairwise nonbonded interactions, which are defined as

$$\Phi^{nb} = \sum_{i,j} [s_{ij}^{es} \frac{q_i q_j}{\epsilon R_{ij}} + s_{ij}^{vdW} (\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6})] \quad (1.2)$$

where the summation is for all atom pairs in the system. The first and second terms in the

square brackets account for the electrostatic and van der Waals (vdW) interactions, respectively. s_{ij}^{es} and s_{ij}^{vdW} are their respective scale factors. For 1-2, 1-3 atom pairs, these interactions are both scaled down to zero; however, for 1-4 atom pairs, the electrostatic interaction is scaled down by $\frac{1.0}{1.2}$, and the vdW interaction is scaled down by 0.5.

MM-based simulation first requires a reliable set of force field parameters. Among all the parameters aforementioned, force constants (k values) are of particular importance because they account for the stiffness of bonded interactions described by the force field. Many sets of parameters have been developed in the past decades for biological and organic molecules and have proven highly useful in simulations. For example, several AMBER-type force fields are available for simulations of these systems. However, with the ongoing development of new molecules that are not described by existing parameters, researchers often encounter a problem in which a certain parameter is missing. Consequently, the development of efficient and reliable protocols for force field parameterization has been and will continue to be an important goal of research in computational chemistry.

1.2 Previous Works

Several numerical optimization algorithms have been used for force field parameterization, such as a genetic algorithm (GA), simplex algorithm, and Monte Carlo techniques. A GA was used by Wang and Kollman to derive force constants for organic molecules.¹⁶ Ryde and coworkers derived parameters for cytochrome P450 and zinc-containing complexes using modified simplex and Newton-Raphson algorithms^{17,18}. Monte Carlo techniques use a random-walk process and were applied to determine force constants for a Cu(I)-ligand complex by Hæffner et al.¹⁹ More recently, Betz and Walker combined GA and simplex methods and implemented them in the Paramfit program.²⁰ These numerical algorithms can usually produce very reliable parameters in terms of reproducing QM-calculated quantities. Despite their usefulness, these numerical optimizations usually require a large number of iterations. As a result, the parameterization processes can be very time-consuming. Moreover, if the number of unknown parameters increases, the dimension of search-space would also increase, which would result in more difficulties to search the global minimum. The high computational demand usually prohibits the use of these methods for complex situations.

A much simpler and faster approach is to take force constants directly from the diagonal elements of the QM-calculated Hessian matrix in internal coordinates. However, there may be many possible choices of internal coordinates, and the extracted diagonal terms usually

vary for the different internal coordinates. To solve this problem, Seminario proposed a method to determine force constants by diagonalizing some 3×3 partial Hessian matrices in Cartesian coordinates.²¹ The Seminario method allows fast parameterization and has been widely used in many studies, especially for deriving force constants required for metalloproteins^{18,20,23–27}, nanosystems^{28,29}, a MOF³⁰, an ionic liquid³¹ and so on. Burger et al. also developed a single projection method to project the Hessian into every single internal coordinate to determine the relevant force constant.³²

In this thesis, we presented some new efficient methods to determine force constant parameters for force fields. The basic idea of our methods is that the MM-calculated Hessian matrix should reproduce the QM-calculated one. Although we consider these methods should be applicable to many different force fields, we here examined them for the AMBER-type force fields. The Seminario method was also tested for comparison, and our methods exhibited better overall performance, in terms of how well the QM-calculated vibrational frequencies were reproduced. We also noticed that just taking equilibrium values from QM-optimized geometries would lead to different geometries after MM-optimization, regardless whether the force constants are determined by our methods or the Seminario method. This problem could be significant when there are large atomic charges and soft bonds in the system. A simple scheme was also developed to overcome this problem by amending equilibrium values.

2. Methods

2.1 Hessian Fitting

In this section, the three methods are introduced to determine force constant parameters. They share some same ideas which are presented below in this preface. The equilibrium values of bond-stretching and angle-bending terms are taken from the QM-optimized geometry. The periodicity and phase parameters of dihedral terms are taken from the general AMBER force field (GAFF). Hence, the force constants are the only unknown parameters. All processes are done at the QM-optimized geometries.

Hessian is a square matrix of second-order partial derivatives of the energy. From the energy expression Eq.(1.1), taking second derivatives with respect to two Cartesian coordinates x_i and x_j , gives:

$$\begin{aligned}
\frac{\partial V^{MM}}{\partial x_i \partial x_j} = & \sum_{bonds} k_b \frac{\partial}{\partial x_i \partial x_j} (r - r_{eq})^2 + \sum_{angles} k_a \frac{\partial}{\partial x_i \partial x_j} (\theta - \theta_{eq})^2 \\
& + \sum_{dihedrals} k_d \frac{\partial}{\partial x_i \partial x_j} \cos(n\phi - \gamma) \\
& + \sum_{impropers} k_{im} \frac{\partial}{\partial x_i \partial x_j} \cos(n\phi - \gamma) + H^{nb}
\end{aligned} \tag{2.1}$$

Apparently, the force constants can be extracted out from the differentiation. Let us simplify (2.1) by defining h :

$$\frac{\partial V^{MM}}{\partial x_i \partial x_j} = \sum_{bonds} k_b h^b + \sum_{angles} k_a h^a + \sum_{dihedrals} k_d h^d + \sum_{impropers} k_{im} h^{im} + H^{nb} \tag{2.2}$$

where

$$\begin{aligned}
h^b &= \frac{\partial}{\partial x_i \partial x_j} (r - r_{eq})^2 \\
h^a &= \frac{\partial}{\partial x_i \partial x_j} (\theta - \theta_{eq})^2 \\
h^d, h^{im} &= \frac{\partial}{\partial x_i \partial x_j} \cos(n\phi - \gamma)
\end{aligned} \tag{2.3}$$

represent the differentiation terms apart from force constants for bond, angle, dihedral or improper terms, respectively. Therefore, these h perform second-order differentiation while force constants remain unchanged. Apparently, h values are only dependent on geometries. Since the whole processes are done at a single geometry, h values are just constants.

For all permutations of Cartesian coordinate x_i and x_j , second derivatives form a symmetric Hessian matrix, in which the off-diagonal elements are duplicated. In fact, most of the programs only provide the lower triangle of Hessian matrices but not the whole ones. Therefore, we only consider the lower triangle of Hessian matrices. Let n be the number of lower triangle elements ($n = \frac{9N^2 + 3N}{2}$, where N is the number of atoms) and just put

these elements in vectors. Eq.(2.2) is transformed to

$$\mathbf{H}^{MM} = \sum_{bonds} k_b \mathbf{h}^b + \sum_{angles} k_a \mathbf{h}^a + \sum_{dihedrals} k_d \mathbf{h}^d + \sum_{impropers} k_{im} \mathbf{h}^{im} + \mathbf{H}^{nb} \tag{2.4}$$

in which the boldface $\mathbf{H}$ and $\mathbf{h}$ are $n \times 1$ column vectors containing all unrepeated second derivatives, compared with the component-wise Eq.(2.2).

Eq.(2.4) indicates that the MM-calculated Hessian $\mathbf{H}^{MM}$ is a linear combination of the $\mathbf{H}_{nb}$, which is the contribution from nonbonded terms, and several $\mathbf{h}$ vectors which account for bonded terms' contributions. The force constants are the coefficients.

To determine each $\mathbf{h}$ vector, an artificial MM Hessian calculation is performed. Its corresponding force constant is set to 1.0 whereas other force constants and all nonbonded

parameters are set to zero. This setting makes the energy expression Eq.(2.1) only have one term and hence the resulted Hessian is just the $\mathbf{h}$ we need. Similarly, to determine the $\mathbf{H}^{nb}$, an artificial MM Hessian calculation is performed with all force constants set to zero but nonbonded parameters unchanged. By this setting, all bonded terms are removed and the resulted Hessian is the $\mathbf{H}^{nb}$ we need.

In total, the number of needed Hessian calculations for determining $\mathbf{h}$ vectors is same to the number of bonded terms and also equal to the number of all bonds, angles, and dihedrals. This is the main computational cost of our methods because other procedures can finish very quickly.

2.1.1 Full Hessian Fitting (FHF)

To reproduce the QM-calculated Hessian matrix, the straightforward way is to let the MM Hessian equal to the QM one, and then solve the force constants from the linear system.

Let the MM Hessian equal to the QM Hessian:

$$\mathbf{H}^{MM} = \sum_{bonds} k_b \mathbf{h}^b + \sum_{angles} k_a \mathbf{h}^a + \sum_{dihedrals} k_d \mathbf{h}^d + \sum_{impropers} k_{im} \mathbf{h}^{im} + \mathbf{H}^{nb} = \mathbf{H}^{QM} \quad (2.5)$$

Moving the nonbonded term to QM side leads to:

$$\sum_{bonds} k_b \mathbf{h}^b + \sum_{angles} k_a \mathbf{h}^a + \sum_{dihedrals} k_d \mathbf{h}^d + \sum_{impropers} k_{im} \mathbf{h}^{im} = \mathbf{H}^{QM} - \mathbf{H}^{nb} \quad (2.6)$$

Now let us not distinguish the type of bonded terms and suppose there are m bonded terms in total. Eq.(2.6) is a linear system, which can be written in the matrix form:

$$\begin{bmatrix} \mathbf{h}^1 & \dots & \mathbf{h}^m \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix} = \mathbf{H}^{QM} - \mathbf{H}^{nb} \quad (2.7)$$

in which $k_1 - k_m$ and $\mathbf{h}^1 - \mathbf{h}^m$ correspond to all m bonded terms. The first term in Eq.(2.7) is an $n \times m$ matrix (n is the number of lower triangle elements) and the second term is an $m \times 1$ vector. Their product yields the right side $n \times 1$ vector.

In almost all cases, the number of unrepeated Hessian elements n is much larger than the number of bonded terms m . Thus, Eq.(2.7) is an overdetermined system which does not have an exact solution in most cases. Rather, an approximate solution can be found by the method of ordinary least squares, which minimizes the error

$$\min \left\| \begin{bmatrix} \mathbf{h}^1 & \dots & \mathbf{h}^m \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix} - (\mathbf{H}^{QM} - \mathbf{H}^{nb}) \right\| \quad (2.8)$$

It is essentially equivalent to:

$$\min \|\mathbf{H}^{QM} - \mathbf{H}^{MM}\| \quad (2.9)$$

In other words, solving Eq.(2.7) by the least-squares method produces a solution that could minimize the differences between QM- and MM-calculated Hessian for all Hessian elements. This approach is therefore called the Full Hessian Fitting (FHF) method.

Although FHF can minimize the error of Hessian in the least-squares sense, this simple approach has a shortcoming for yielding force constants. When minimizing the quantity (2.9) by the ordinary-least-squares method, compromises are made for all the equations in the linear system to minimize the overall root-mean-square error (RMSE). This procedure implies that all Hessian elements are equal. However, as will be shown in the next section, Hessian elements have different importance for different force constants. Minimizing overall RMSE may lead to large relative errors for elements whose absolute value is small. Consequently, when there is a relatively weak interaction, such as the case of dihedral terms, the determined force constants may not be good.

2.1.2 Partial Hessian Fitting (PHF)

FHF treats all Hessian elements equally and solves them together in the least-squares sense. Nevertheless, some specific Hessian elements are of particular importance for certain force constants. This inequality should be considered in parameterization.

Here we introduce the Partial Hessian Fitting (PHF) method that focuses on some 3×3 partial Hessian matrices. The selected partial Hessian matrices are corresponding to two terminal atoms of dihedrals, angles, or bonds. In other words, each of them contains 9 second derivatives with respect to the Cartesian coordinates of two terminal atoms. Importantly, most of the terms in Eq.(2.4) vanish in these partial Hessian.

Let us now consider a chain segment within a molecule, as shown in Fig 2.1.

Figure 2.1 Schematic illustration of a chain segment within a molecule

For the dihedral $i-j-k-l$, angle $i-j-k$ and bond $i-j$, the selected partial Hessian matrices are $\mathbf{H}_{il}^{MM}$, $\mathbf{H}_{ik}^{MM}$ and $\mathbf{H}_{ij}^{MM}$, respectively. For these partial Hessian matrices, only some closely related terms remain.

Dihedral term

For the partial Hessian of the atom pair $i-l$, only the dihedral term for the dihedral angle $i-j-k-l$ and the nonbonded term between two atoms i and l remain:

$$\mathbf{H}_{il}^{MM} = k_d \mathbf{h}_{il}^d(i-j-k-l) + \mathbf{H}_{il}^{nb}(i, l) \quad (2.10)$$

The other terms vanish since they do not contain the coordinates of the two atoms i and j simultaneously. Similar to Eq.(2.5)–(2.7), we let it equal to the QM Hessian in the same region, and then solve the linear system:

$$k_d \mathbf{h}_{il}^d(i-j-k-l) = \mathbf{H}_{il}^{QM} - \mathbf{H}_{il}^{nb}(i, l) \quad (2.11)$$

which is actually 9 equations for the 9 elements in the selected partial Hessian. Same to Eq.(2.8) and (2.9), the ordinary-least-squares method can produce an approximate solution which achieves the minimum requirement of

$$\min \left\| \mathbf{H}_{il}^{QM} - \mathbf{H}_{il}^{MM} \right\| \quad (2.12)$$

By this approach, all force constants for dihedral terms can be determined.

Angle-bending term

For the partial Hessian of the atom pair $i-k$, besides the angle term for the angle $i-j-k$, a few dihedral terms containing the two atoms i and k also remain, which are denoted by the summation term in Eq.(2.13):

$$\mathbf{H}_{ik}^{MM} = k_a \mathbf{h}_{ik}^a(i-j-k) + \sum_{\text{dihedrals}} k_d \mathbf{h}_{ik}^d(i, k, \dots) \quad (2.13)$$

In the case of Fig 2.1, they are the dihedrals $i-j-k-l$, $i'-j-k-l$ and $i''-j-k-l$. Equaling the MM Hessian and the QM one leads to the linear system:

$$k_a \mathbf{h}_{ik}^a(i-j-k) = \mathbf{H}_{ik}^{QM} - \sum_{\text{dihedrals}} k_d \mathbf{h}_{ik}^d(i, k, \dots) \quad (2.14)$$

In the previous step, the force constants of dihedral terms are already determined. Therefore, Eq.(2.14) can also be treated with the ordinary-least-squares method to produce all the force constants for angle-bending terms.

Bond-stretching term

Similarly, for the partial Hessian of the atom pair $i-j$, besides the bond term for the bond $i-j$, some angle and dihedral terms containing the two atoms i and j also remain.

$$\mathbf{H}_{ij}^{MM} = k_b \mathbf{h}_{ij}^b(i-j) + \sum_{\text{angles}} k_a \mathbf{h}_{ij}^a(i, j, \dots) + \sum_{\text{dihedrals}} k_d \mathbf{h}_{ij}^d(i, j, \dots) \quad (2.15)$$

The linear system is

$$k_b \mathbf{h}_{ij}^b(i-j) = \mathbf{H}_{ij}^{QM} - \sum_{\text{angles}} k_a \mathbf{h}_{ij}^a(i, j, \dots) - \sum_{\text{dihedrals}} k_d \mathbf{h}_{ij}^d(i, j, \dots) \quad (2.16)$$

from which the force constants of bond-stretching terms can be solved, because angles and dihedrals are already determined before.

Ring cases and Improper term

However, for 6- or 5-membered ring cases, or for the case where an improper term is defined (Fig 2.2), the procedure is slightly different.

Figure 2.2 a) Two dihedral angles including atoms i and l in a six-membered ring. b) A dihedral angle $i-j-k-l$ and an angle $i-j'-l$ in a five-membered ring. c) An improper angle $i-l-j-k$ and an angle $i-j-k$

For the six-membered ring in Fig 2.2 (a), the Hessian of the atom pair $i-l$ is

$$\mathbf{H}_{il}^{MM} = k_d \mathbf{h}_{il}^d(i-j-k-l) + k'_d \mathbf{h}_{il}^{d'}(i-j'-k'-l) + \mathbf{H}_{il}^{nb} \quad (2.17)$$

which contain two dihedral terms simultaneously.

For the five-membered ring in Fig 2.2 (b), the Hessian of the atom pair $i-l$ is

$$\mathbf{H}_{il}^{MM} = k_d \mathbf{h}_{il}^d(i-j-k-l) + k_a \mathbf{h}_{il}^a(i-j'-l) + \sum_{\text{dihedrals}} k_d \mathbf{h}_{il}^{d'}(i, l, \dots) \quad (2.18)$$

which contain a dihedral term and also an angle term.

Also, when an improper term is defined, as in Fig 2.2 (c), the Hessian of the atom pair $i-k$ is

$$\mathbf{H}_{ik}^{MM} = k_a \mathbf{h}_{ik}^a(i-j-k) + k_{im} \mathbf{h}_{ik}^{im}(i-l-j-k) + \sum_{\text{dihedrals}} k_d \mathbf{h}_{ik}^d(i, k, \dots) \quad (2.19)$$

which contain an angle term and also an improper term.

The last terms in Eq.(2.17)–(2.19) are either known or already determined in the former steps. For these linear systems, there are two unknown force constants and 9 equations. Still, the ordinary-least-squares method can be applied to determine two force constants simultaneously in the same manner and minimizes the error between the QM- and MM-calculated Hessian in the 3×3 regions.

Limitation in 3- and 4-membered ring cases

As described above, PHF first determines the force constants for dihedral terms, and then for angle terms, and at last for bond terms. This special treatment is to remove the dependence of Cartesian Hessian by successively subtracting the contributions from former terms. As discussed above, there are still some coupling between different terms in the ring cases. For 6- and 5-membered ring cases, we have shown PHF is still applicable. However,

for 3- and 4-membered ring cases, because more than 2 terms are coupled with each other, it is not possible to determine the first force constant and subsequently eliminate the dependence. Hence, PHF has the limitation in 3- and 4-membered ring cases, although this situation does not usually occur.

2.1.3 Internal Hessian Fitting (IHF)

Instead of using Cartesian Hessian, conventionally, force constants were extracted directly from diagonal elements of the Hessian in internal coordinates (we call it *internal Hessian* hereafter).^{33–36} However, this method is very sensitive to the choice of internal coordinates. The extracted force constants can vary a lot for different choices of internal coordinate.^{21,37} To remove this arbitrariness, Burger et al. proposed a single projection method to project the Cartesian Hessian into every single internal coordinate.³² However, because the internal coordinates are not independent, the single projection method uses an iterative algorithm starting from an initial guess to reproduce the QM internal Hessian. In addition, each force constant is determined separately. Thus, the computational cost is still relatively high.

Here we introduce the Internal Hessian Fitting (IHF) method to determine force constants using internal Hessian. Similar to FHF and PHF, this method does not need an iterative algorithm and can be done analytically. Importantly, all force constants in the system are determined simultaneously but not separately to consider the overall dependence of internal coordinates. With much lower computational demand, the diagonal elements of Hessian in redundant internal coordinates can be reproduced. The redundant internal coordinates are defined for all bonded terms. In other words, all bond distances, bonded angles, bonded dihedrals, and improper angles are defined as the redundant internal coordinates. This exclusive definition removes the arbitrariness of the choice of internal coordinates. More importantly, this is the most natural coordinate system for the force field because it corresponds to all bonded terms in Eq.(1.1). If we use one Fourier term for dihedrals, the force constants and the internal coordinates would have a one-to-one correspondence.

Conventionally, the internal Hessian is not calculated directly but is transformed from the Cartesian Hessian using the Wilson B matrix³⁸. Despite the convention, directly taking energy derivatives with respect to internal coordinates should also be correct in principle. The Eq.(2.1) is still correct for internal Hessian just by changing the x to q .

$$\begin{aligned}
\frac{\partial V^{MM}}{\partial q_i \partial q_j} = & \sum_{bonds} k_b \frac{\partial}{\partial q_i \partial q_j} (r - r_{eq})^2 + \sum_{angles} k_a \frac{\partial}{\partial q_i \partial q_j} (\theta - \theta_{eq})^2 \\
& + \sum_{dihedrals} k_d \frac{\partial}{\partial q_i \partial q_j} \cos(n\phi - \gamma) \\
& + \sum_{impropers} k_{im} \frac{\partial}{\partial q_i \partial q_j} \cos(n\phi - \gamma) + H^{nb}
\end{aligned} \tag{2.20}$$

in which the differentiation terms are actually dependent on the choice of internal coordinates and cannot be calculated directly. However, the same artificial MM calculations described in the preface of Section 2.1 can be applied to determine these values. All derivations in Section 2.1 and 2.1.1 still hold true for the internal Hessian just by changing the symbols. The internal Hessian is still a linear combination of the $\mathbf{H}_q^{nb}$ and $\mathbf{h}_q$ vectors:

$$\mathbf{H}_q^{MM} = \sum_{bonds} k_b \mathbf{h}_q^b + \sum_{angles} k_a \mathbf{h}_q^a + \sum_{dihedrals} k_d \mathbf{h}_q^d + \sum_{impropers} k_{im} \mathbf{h}_q^{im} + \mathbf{H}_q^{nb} \tag{2.21}$$

which is corresponding to Eq.(2.4). Finally, we arrive at the same step corresponding to Eq.(2.6).

$$\begin{bmatrix} \mathbf{h}_q^1 & \cdots & \mathbf{h}_q^m \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix} = \mathbf{H}_q^{QM} - \mathbf{H}_q^{nb} \tag{2.22}$$

In FHF, we use all Cartesian Hessian elements and solve an overdetermined system because the importance of Cartesian Hessian elements is unclear. In the contrary, the diagonal elements of internal Hessian are of particular importance for determining force constants of quadratic functions. From Eq.(2.22), we just pick out the equations for diagonal elements and they form a new linear system.

$$\begin{bmatrix} \mathbf{h}'_q^1 & \cdots & \mathbf{h}'_q^m \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix} = \mathbf{H}'_q^{QM} - \mathbf{H}'_q^{nb} \tag{2.23}$$

where the $\mathbf{h}'$ vectors only have diagonal elements of the $\mathbf{h}$ vectors. Because the internal coordinates are defined one-to-one corresponding to the force constants, the number of unknowns and the number of equations are same. Hence, Eq.(2.23) can be solved exactly. In other words, if we do not perform MM optimization and just calculate the internal Hessian using the IHF parameters at the same geometry, the diagonal elements could be exactly reproduced.

Because of the dependence, a diagonal element is not only contributed by its corresponding bonded term, but also by terms corresponding to other internal coordinates. In the IHF method, the dependence of internal coordinates is considered by including the non-corresponding diagonal elements in $\mathbf{h}_q$.

2.2 Amendment of Equilibrium Values (Katachi)

If the equilibrium values for bond-stretching and angle-bending terms (r_{eq} and θ_{eq} in Eq.(1.1)) are taken from the QM-optimized geometry, the QM-optimized geometry would be the minimum for these terms. For dihedral terms, they can also be parameterized to meet the minimum by fitting to a dihedral scan with more Fourier terms. Thus, if we look at the first derivatives (forces), bonded terms would have zero contribution. Nevertheless, nonbonded terms would inevitably have nonzero contributions to the forces, as long as nonbonded interactions exist (they are removed for 1-2 and 1-3 atom pairs). As the result, the QM-optimized geometry would not be the minimum for the force field. Consequently, after the MM optimization, the geometry would change. This change is usually negligible because the bonded terms are very strong and can resist the nonbonded interactions. However, sometimes the nonbonded interactions could overwhelm the bonded interactions, especially in the cases of soft bonds and large atomic charges within a molecule, thereby resulting in a large deviation between QM- and MM-optimized geometries.

To remedy this issue, we proposed a simple algorithm to amend the equilibrium values in Eq.(1.1) and to minimize the deviation between QM- and MM-optimized geometries.

- a. Run MM optimization to get MM-optimized bond distances and angles $X_i(MM)$ and find the deviation $\Delta_i = X_i(QM) - X_i(MM)$ for every bond and angle.
- b. If all Δ_i values are smaller than a given threshold (here we use 0.0001 Å for bond distances and 0.002° for angles), STOP. Otherwise, all r_{eq} and θ_{eq} are incremented by their corresponding Δ_i and then GOTO (a).

With this simple procedure, we expect that the MM-optimized geometries could be improved to resemble the QM-optimized ones. We shall henceforth call this procedure *Katachi amendment* (K).

3. Implementation

3.1 Sample molecules, QM and MM Calculations

In this study, all the QM and MM calculations were performed using Gaussian 09 software.³⁹ We examined 20 small molecules (**1—20**) and also three representative secondary building units of MOFs (MOF-SBUs) (**21—23**). These sample molecules are shown in Fig 3.1.

The reference QM data including the optimized geometries and Hessian were calculated using density functional theory at the B3LYP/6-31+G(d) (5d) level^{40—44}; but for the Cu and Zn atoms in the MOF-SBUs, the SDD effective core potential basis set was

used⁴⁵. For the QM optimization, keywords *opt=(verytight, maxstep=7, notrust)* and *int=ultrafine* were used and the latter one was also used for the Hessian calculation. MM frequency calculation was done by *opt=(calcall, z-matrix, verytight)* keyword to obtain the precise minima. For most of the sample molecules, the close-shell singlet spin state was used; however, NO₂ (**4**) was assumed to have doublet spin state, and for Cu-MOF-SBU (**21**), the open-shell singlet ground state was used.

For nonbonded terms in Eq.(1.2), RESP charges were derived using Gaussian 09 (with *pop=mk iop(6/33=2,6/42=17,6/41=10) int=ultrafine* keywords) and AmberTools (with *-c resp* argument)^{39,46}. For most of the atoms, vdW parameters were taken from the GAFF force field¹⁵, but for Cu, Zn, Al and B, whose parameters are not available in GAFF, they were taken from other papers by Li et al.^{47,48} and A. Tafi et al.⁴⁹, respectively. These vdW parameters were also used in the charge calculation using Gaussian 09. For comparison, we also determined bond-stretching and angle-bending parameters using the Seminario method. Note that the Seminario method cannot parameterize the dihedral terms used in AMBER force field; hence, the comparison was done by using dihedral parameters determined by PHF.

Usually, there are many equivalent internal coordinates within a molecule. For example, benzene has six equivalent C-C bonds and C-H bonds; all C-C-H angles are also equivalent. These equivalent internal coordinates are separately considered and parameterized, but their parameters are finally averaged to find a representative value. Sometimes, although some dihedrals are not equivalent, they are treated equally by using wildcards. For example, in benzene, there are three types of dihedrals: H-C-C-C, H-C-C-H and C-C-C-C; in GAFF, they are equally treated as the X-C-C-X type, where X is the wildcard. In such cases, the parameters of these “quasi”-equivalent dihedrals are also averaged to find the final representative parameters.

Figure 3.1 Sample molecules considered in this study

3.2 Evaluating Performance

To access the performance, we compared the QM- and MM-calculated vibrational frequencies. The following L₁ norm of the deviations between the QM- and MM-calculated harmonic frequencies was used:

$$\frac{\Delta_{\text{freq}}}{N_q} = \sum_i^{N_q} |\tilde{\nu}_i^{QM} - \tilde{\nu}_{j[i]}^{MM}| / N_q \quad (3.1)$$

where $\tilde{\nu}_i^{QM}$ and $\tilde{\nu}_{j[i]}^{MM}$ are the QM-calculated wavenumbers and the corresponding MM-calculated ones, obtained at their individually optimized geometries. It should be noted that the wavenumbers are sorted in ascending orders; thus, the sequences of QM- and MM-calculated normal modes are usually not same. This mismatch should be rectified before comparing the deviation between the QM- and MM-calculated wavenumbers. To rectify it, we considered the displacement vector $\mathbf{I}_i^{QM}$ and $\mathbf{I}_j^{MM}$ for QM- and MM-calculated normal modes. To judge the similarity of two normal modes, we defined the following similarity

index (SI_{ij}):

$$SI_{ij} = |\cos \theta_{ij}| \quad (3.2)$$

where θ_{ij} is the angle between $\mathbf{I}_i^{QM}$ and $\mathbf{I}_j^{MM}$. Each i -th QM normal mode was compared with every MM normal mode, starting from the high-frequency modes. If the j -th MM normal mode gave largest SI_{ij} among all MM normal modes, we considered it represents the same vibration as the i -th QM normal mode. Using this method, we found the MM counterparts of all QM-calculated frequencies, and then the quantity in Eq.(3.1) was used to evaluate the performance.

4. Results and Discussion

4.1 Test of Methods

The frequency data for H_2O_2 are summarized in Table 4.1. FHF, PHF, and IHF provided

$\frac{\Delta_{freq}}{N_q}$ values of 98.6, 39.3, and 18.8 cm^{-1} , respectively, whereas the Seminario method

provided 39.1 cm^{-1} with dihedral parameters taken from PHF. Thus, the performance of PHF was similar to that of the Seminario method for this molecule, but IHF was much better than all others. Burger et al. used single projection method to determine the parameters for

this molecule at the HF/6-31G(d) level, and the provided $\frac{\Delta_{freq}}{N_q}$ value was 43.2 cm^{-1} . We

also determined the parameters for this molecule at the same level, and the resulted $\frac{\Delta_{freq}}{N_q}$

were 112.6, 40.8, and 19.6 cm^{-1} for FHF, PHF, and IHF, respectively. PHF and IHF were also better in this case. As described above, our methods follow the very simple routes that do not require a dihedral scan calculation when determining the force constant for a dihedral angle. Despite this simplicity, for this simple case, the low-frequency mode associated with a dihedral torsion (379.9 cm^{-1}) was well reproduced by the PHF and IHF parameters (400.7 and 401.2 cm^{-1} , respectively). Although, for more complicated cases, the dihedrals may not be easy to parameterize, which would be explained at the end of this section.

Table 4.1 Frequencies (cm^{-1}) for hydrogen peroxide calculated using different methods and their deviations from QM frequencies (Δ_{freq} and $\frac{\Delta_{freq}}{N_q}$, in cm^{-1}).

Freq. no.	B3LYP/ 6-31+G(d)	FHF	PHF	IHF	Seminario*
1	379.9	531.1	400.7	401.2	401.2
2	948.0	871.6	862.2	963.0	865.1
3	1321.4	1217.8	1414.0	1386.2	1323.8
4	1463.4	1280.0	1478.6	1461.7	1389.6
5	3706.8	3750.4	3722.5	3710.0	3739.2
6	3707.1	3740.5	3712.5	3699.8	3729.3
Δ_{freq}	-	591.5	235.5	113.1	234.8
Δ_{freq}/N_q	-	98.6	39.3	18.8	39.1

* The k_d value was taken from PHF.

Table A1 (shown in the Appendices) summarizes the deviation between QM- and MM-calculated frequencies using the FHF, PHF, and IHF parameter sets for molecules **1—20**. For comparison, the MM data produced by the Seminario method were also listed. For most of the small molecules **1—20**, PHF and IHF gave overall the smaller $\frac{\Delta_{\text{freq}}}{N_q}$ values than those of the Seminario method. The only exceptions were the cases of NO₂ and H₂O₂, where the differences are very small. From the sum, mean and root mean square (RMS) values of $\frac{\Delta_{\text{freq}}}{N_q}$ shown in Table A1, it is clear that our methods have better overall performance, and this advantage is more significant in larger molecules. For example, for C₆H₆, FHF, PHF and IHF provided $\frac{\Delta_{\text{freq}}}{N_q}$ values of 58.5, 79.0 and 51.2 cm⁻¹, whereas the Seminario parameters provided that of 165.0 cm⁻¹.

The three MOF-SBUs **21—23** were then used to examine the performance of different methods. FHF was not shown here because it did not perform well for the MOF-SBUs. We expect IHF would also perform well for the MOF-SBUs, but unfortunately, the program for IHF was not fully completed due to the limited time to support MOF-SBUs. Therefore, here we show the performance of PHF and the Seminario method to parameterize three MOF-SBUs in Table 4.2. These data were obtained with $s_{ij}^{es} = 0.0$ for 1-4 electrostatic

interactions because the geometries distorted significantly with the original setting. Again, the PHF parameters performed better than the Seminario parameters, where the mean values of deviation were 115.0 and 145.0 cm⁻¹, respectively.

Table 4.2 Comparison of Δ_{freq} (cm⁻¹), $\frac{\Delta_{freq}}{N_q} \Delta_{freq}/N_q$ (cm⁻¹), and rmsd values (Å) for MOF-SBUs 21–23 Obtained with the PHF and Seminario Methods.

SN	Molecule	PHF			Seminario*		
		Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd
21	Cu(II) ₂ (HCOO) ₄	5114.5	106.6	0.202	5185.9	108.0	0.185
22	Al(III) ₃ O(HCOO) ₆ ⁺	12376.8	158.7	0.336	15552.6	199.4	0.319
23	Zn(III) ₄ O(HCOO) ₆	6455.9	79.7	0.351	10325.0	127.5	0.341
	Sum		344.9	0.889		434.9	0.845
	Mean		115.0	0.296		145.0	0.282
	RMS		119.6	0.304		150.2	0.290

* The k_d value was taken from PHF.

4.2 Test of Katachi Amendment

Although the force constants were successfully determined and our methods have the better performance compared with the Seminario method, the rmsd values were usually quite large, as shown in Table A1 and 4.2. Simple molecules **1**–**4** (HF, CO, H₂O, and NO₂) and **6**–**8** (NH₃, CH₄, and SiH₄) had zero rmsd values because they neither have dihedral angles nor have nonbonded interactions. The three C-H or Si-H hydrogen atoms in molecules **12**–**13** (CH₃OH and SiH₃OH) are not exactly equivalent as well as the related bonds and angles, but we decided to treat them as equivalent in the force field. Therefore, they had large rmsd values but did not mean the poor performance of our methods. The dihedral parameters obtained by all methods were not good to describe the dihedral angles in molecules **11** (NH₂NH₂), **19** (CH₃-NHCO-CH₃), and **20** (C₆H₅-C₆H₅), and thus their large rmsd values were mainly due to the inaccurate description of dihedrals. More Fourier terms are usually needed to describe these complex dihedral angles, which will be discussed at the end of this section.

For other molecules **5**, **9**, **10**, **14**–**18**, and especially for the MOF-SBUs **21**–**23**, their rmsd values were mainly due to the distortion problem described in Section 2.2. Figure 4.1 shows the superposition of QM- and MM(PHF)-optimized geometries for the three MOF-SBUs, and the deviations are seen to very large. In general, coordinate bonds in MOF-SBUs are not as strong as covalent bonds in organic molecules; but the partial atomic charges on metal atoms could be relatively large. Therefore, the electrostatic interaction distorted the geometries significantly. We attempted to remove the 1–4 electrostatic

interactions within MOF-SBUs using $s_{ij}^{es} = 0.0$, but the rmsd values were still large, as shown in Table 4.2. Hence, we applied Katachi amendment to remedy this distortion problem.

Figure 4.1 Superposition of the QM (red) and MM (yellow) optimized geometries of a) **21**, b) **22**, and c) **23**. Here, the force constant parameters for MM (with $s_{ij}^{es} = 0.0$ for 1–4 interactions) were determined by the PHF method.

Table A2 summarizes the data for small molecules that had the distortion problem using different methods with Katachi amendment. The rmsd values clearly show that Katachi amendment improved the geometries and most of the rmsd values were improved close to 0.000. The mean values of rmsd were 0.006, 0.000, 0.000, and 0.001 for FHF-K, PHF-K, IHF-K, and Seminario-K, respectively.

Katachi amendment was also applied to the three MOF-SBUs and the results are shown in Table 4.3. The rmsd values were also improved significantly compared with those obtained without Katachi amendment. The mean rmsd values were improved from 0.296&0.282 to 0.009&0.008 for PHF and Seminario, respectively. This improvement can also be seen from the superposition of geometries. In Figure 4.1, we can see the severe deviation between the QM- and MM(PHF)-optimized geometries for the three MOF-SBUs. With Katachi amendment, the MM(PHF-K)-optimized geometries are seen to agree very well with the QM geometries, as shown in Figure 4.2

Table 4.3 Comparison of Δ_{freq} (cm^{-1}), $\frac{\Delta_{freq}}{N_q}$ (cm^{-1}), and rmsd values ($\text{\AA}$) for

MOF-SBUs 21–23 Obtained with the PHF and Seminario Methods with Katachi Amendment

SN	Molecule	PHF-K			Seminario-K*		
		Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd
21	$\text{Cu(II)}_2(\text{HCOO})_4$	4623.8	96.3	0.000	5707.1	118.9	0.000
22	$\text{Al(III)}_3\text{O}(\text{HCOO})_6^+$	9262.4	118.7	0.016	11896.5	152.5	0.012
23	$\text{Zn(III)}_4\text{O}(\text{HCOO})_6$	6327.2	78.1	0.011	6481.6	80.0	0.011
	Sum		293.1	0.027		351.4	0.023

	Mean		97.7	0.009		117.1	0.008
	RMS		99.1	0.011		120.8	0.009

* The k_d value was taken from PHF.

Figure 4.2 Superposition of the QM (red) and MM (yellow) optimized geometries of a) 21, b) 22, and c) 23. Here, the force constant parameters and the equilibrium bond distances and angles for MM (with $s_{ij}^{es} = 0.0$ for 1–4 interactions) were determined by the PHF-K method.

4.3 Discussion

Figure 4.3 summarizes $\frac{\Delta_{\text{freq}}}{N_q}$ values for all molecules **1—23**, with Katachi amendment if applicable. The figure clearly shows the overall better performance of the PHF and IHF methods compared with the Seminario method, in terms of how well the QM-calculated frequencies were reproduced. The sum, mean, RMS values of $\frac{\Delta_{\text{freq}}}{N_q}$ also highlighted the better performance of our methods, as shown in Table A1 and 4.2.

Figure 4.3. Comparison of the $\frac{\Delta_{freq}}{N_q}$ values for molecules **1–23**. MM data were calculated using different methods; for molecules **5, 9, 10, 14–18**, and **21–23**, Katachi amendment was applied.

Apart from the performance, another important advantage of our methods is the flexibility. Unlike the Seminario method and also the single projection method, our methods do not require the energy term to be quadratic. As long as the energy term is linear about the force constant, and hence the Hessian can be represented as the linear combination of individual terms in Eq.(2.4), our methods could be applied. Hence, although here we only examined our method for the AMBER-type force field, these methods should be also applicable for force fields containing other types of terms such as the bond-angle coupling terms $k(r - r_{eq})(\theta - \theta_{eq})$. In fact, we have shown this flexibility by parameterizing dihedral terms as well.

In addition, even if dihedrals are defined quadratic, the Seminario and single projection methods would encounter the dependence problem that the selected Cartesian or internal Hessian is both contributed by nonbonded interactions and the dihedral term, but the nonbonded terms' contribution was not removed. By contrast, in the PHF method, the nonbonded interaction was separated as shown in Eq.(2.11); in IHF, the dependence is considered in a linear system. We believe this kind of consideration is an important feature of our methods.

We also compare key ideas and differences of our methods. Although not fully implemented due to the limited time, the IHF method exhibited the best performance in the tests for molecule **1–20**. The diagonal elements of internal Hessian reflect the force constants of quadratic terms, and they were exactly reproduced by the IHF method using a linear system. Consequently, the IHF method reproduced frequencies very well. The PHF method tends to parameterize each force constant separately, focusing on some 3×3 partial Hessian matrices, to eliminate the dependence of Cartesian Hessian elements by means of successive subtractions. The force constants were solved in the least-squares sense to reproduce the partial Hessian. The PHF method also provided good frequencies due to the consideration of dependence of Cartesian Hessian. The FHF method is the simplest idea to reproduce Cartesian Hessian. Despite the relatively worse performance compared with the IHF or PHF methods, FHF has an advantage that it is generally applicable for any kind of systems, whereas PHF cannot be applied in 3- or 4-membered cases due to the complexity of dependence, and IHF was not fully completed yet.

Finally, we comment on the common limitation about the parameterization of dihedral parameters. Bond-stretching and angle-bending are described using a quadratic

approximation since these interactions are quite strong. For simple cases of dihedral torsion where using one Fourier term is good enough, such as the cases of H_2O_2 and C_6H_6 , our methods also produced good dihedral force constants. However, in many cases, dihedral angles can easily change over full 360° due to low energy barriers; also, the rotation profiles can be complicated. Therefore, as opposite to bond-stretching and angle-bending terms, dihedral terms are described by Fourier series, which in principle could reproduce any kind of periodic profile by an infinite sum. In practice, more than 3 Fourier terms are usually needed to describe a complicated rotation profile in the acceptable precision and a dihedral scan calculation is usually necessary for such parameterization, which is time-consuming. Although our methods can reproduce the Hessian at the equilibrium point, it is not possible to well reproduce the torsional profile from the information of the single geometry. Despite this limitation, the idea of this study has inspired us for further development for parameterizing dihedrals. We expect that including the consideration of Hessian in parameterizing dihedrals can enhance the quality of dihedral parameters and reduce the computational cost.

5. Conclusion

In this thesis, we have presented several methods based on Hessian fitting to determine force constants used in molecular mechanics. The basic idea of our methods is that the QM-calculated Hessian should be reproduced by force fields. We have shown that these methods can be done analytically and have better performance compared with an existing method. We expect that these unique parameterization methods will help researchers in MM simulation.

Appendices

Table A1 Comparison of Δ_{freq} (cm^{-1}), $\frac{\Delta_{freq}}{N_q}$ (cm^{-1}), and rmsd values ($\text{\AA}$) for Molecules 1-20 Obtained with Different Methods.

SN	Molecule	FHF			PHF			IHF			Seminario*		
		Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd
1	HF	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000
2	CO	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000
3	H ₂ O	341.3	113.8	0.000	98.4	32.8	0.000	77.8	25.9	0.000	141.4	47.1	0.000
4	NO ₂	246.3	82.1	0.000	213.7	71.2	0.000	265.4	88.5	0.000	203.4	67.8	0.000
5	H ₂ O ₂	591.5	98.6	0.048	235.5	39.3	0.020	113.1	18.8	0.020	234.8	39.1	0.023
6	NH ₃	627.4	104.6	0.000	434.3	72.4	0.000	270.7	45.1	0.000	497.1	82.8	0.000
7	CH ₄	1124.8	125.0	0.000	550.2	61.1	0.000	250.9	27.9	0.000	572.3	63.6	0.000
8	SiH ₄	391.7	43.5	0.000	452.7	50.3	0.000	291.3	32.4	0.000	664.7	73.9	0.000
9	C ₂ H ₄	919.6	76.6	0.010	871.0	72.6	0.008	684.4	57.0	0.008	1648.2	137.4	0.006
10	BH ₂ NH ₂	1220.2	101.7	0.020	1200.2	100.0	0.017	1177.7	98.1	0.017	2096.4	174.7	0.012
11	NH ₂ NH ₂	2655.7	221.3	0.245	1372.2	114.4	0.543	1195.2	99.6	0.109	1637.7	136.5	0.542
12	CH ₃ OH	1378.3	114.9	0.028	811.1	67.6	0.028	676.7	56.4	0.028	1624.4	135.4	0.028
13	SiH ₃ OH	542.4	45.2	0.088	689.4	57.4	0.066	785.4	65.5	0.071	1551.8	129.3	0.070
14	CH ₃ CH ₃	2045.1	113.6	0.002	1638.8	91.0	0.002	1144.6	63.6	0.002	1970.0	109.4	0.002
15	BH ₃ NH ₃	2046.1	113.7	0.114	1564.8	86.9	0.098	964.3	53.6	0.062	2216.7	123.1	0.096
16	C ₄ H ₆	1418.7	59.1	0.009	1324.0	55.2	0.008	901.2	37.5	0.008	3007.8	125.3	0.006
17	C ₆ H ₆	1755.5	58.5	0.008	2370.0	79.0	0.008	1535.5	51.2	0.007	4950.8	165.0	0.008
18	B ₃ N ₃ H ₆	1810.6	60.4	0.027	2775.1	92.5	0.027	1683.8	56.1	0.023	4895.5	163.2	0.032
19	CH ₃ -NHCO-CH ₃	3171.2	105.7	0.459	2670.1	89.0	0.473	1976.0	65.9	0.478	4692.8	156.4	0.435
20	C ₆ H ₅ -C ₆ H ₅	7826.8	130.4	0.322	5885.6	98.1	0.302	5407.1	90.1	0.534	11214.3	186.9	0.293

	Sum		1768.6			1330.9			1033.2			2116.7	
	Mean		88.4			66.5			51.7			117.6	
	RMS		100.7			73.0			59.0			125.6	

* The k_d value was taken from PHF.

Table A2 Comparison of Δ_{freq} (cm^{-1}), $\frac{\Delta_{freq}}{N_q}$ (cm^{-1}), and rmsd values ($\text{\AA}$) Obtained with Different Methods and Katachi Amendment

SN	Molecule	FHF-K			PHF-K			IHF-K			Seminario-K*		
		Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd	Δ_{freq}	Δ_{freq}/N_q	rmsd
5	H ₂ O ₂	576.1	96.0	0.042	244.6	40.8	0.000	118.4	19.7	0.000	220.4	36.7	0.003
9	C ₂ H ₄	905.1	75.4	0.000	882.7	73.6	0.000	688.7	57.4	0.000	1661.9	138.5	0.000
10	BH ₂ NH ₂	1237.7	103.1	0.000	1185.2	98.8	0.000	1166.8	97.2	0.000	2076.4	173.0	0.000
14	CH ₃ CH ₃	2042.9	113.5	0.000	1637.1	90.9	0.000	1141.2	63.4	0.000	1971.6	109.5	0.000
15	BH ₃ NH ₃	2069.6	115.0	0.000	1473.3	81.8	0.000	985.5	54.7	0.000	2127.0	118.2	0.000
16	C ₄ H ₆	1395.7	58.2	0.003	1335.5	55.6	0.002	905.5	37.7	0.002	3029.6	126.2	0.002
17	C ₆ H ₆	1750.7	58.4	0.000	2381.7	79.4	0.000	1535.5	51.2	0.000	4971.2	165.7	0.000
18	B ₃ N ₃ H ₆	1817.8	60.6	0.000	2730.6	91.0	0.000	3999.5	133.3	0.000	4800.4	160.0	0.000
	Sum		680.2	0.045		612.0	0.002		514.7	0.002		1027.9	0.005
	Mean		85.0	0.006		76.5	0.000		64.3	0.000		128.5	0.001
	RMS		88.1	0.015		78.6	0.001		72.4	0.001		134.8	0.001

* The k_d value was taken from PHF.

References

- [1] U. Burkert, N. L. Allinger, *Molecular Mechanics*; *American Chemical Society: Washington, DC*, **1982**.
- [2] P. A. Kollman, **Chem. Rev.** **1993**, *93*, 2395.
- [3] M. Karplus, J. A. McCammon, *Nat. Struct. Biol.* **2002**, *9*, 646.
- [4] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
- [5] W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem. Int. Ed.* **1990**, *29*, 992.
- [6] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, *Science* **2010**, *330*, 341.
- [7] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474.
- [8] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- [9] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440.
- [10] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 490.
- [11] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765.
- [12] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr. D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- [13] J. Wang, P. Cieplak, P. A. Kollman, *J. Comput. Chem.* **2000**, *21*, 1049.
- [14] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 1999.
- [15] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
- [16] J. M. Wang, P. A. Kollman, *J. Comput. Chem.* **2001**, *22*, 1219.
- [17] P. Rydberg, L. Olsen, P. O. Norrby, U. Ryde, *J. Chem. Theory Comput.* **2007**, *3*, 1765.

- [18] L. Hu, U. Ryde, *J. Chem. Theory Comput.* **2011**, 7, 2452.
- [19] F. Hæffner, T. Brinck, M. Haeberlein, C. J. Moberg, *Mol. Struct. (Theochem)* **1997**, 397, 39.
- [20] R. M. Betz, R. C. Walker, *J. Comput. Chem.* **2015**, 36, 79.
- [21] J. M. Seminario, *Int. J. Quantum Chem.* **1996**, 60, 1271.
- [22] F. Lin, R. Wang, *J. Chem. Theory Comput.* **2010**, 6, 1852.
- [23] K. Nilsson, D. Lecerof, E. Sigfridsson, U. Ryde, *Acta Crystallogr. D* **2003**, 59, 274.
- [24] M. B. Peters, Y. Yang, B. Wang, L. F€usti-Molnar, M. N. Weaver, K. M. Merz, Jr., *J. Chem. Theory Comput.* **2010**, 6, 2935.
- [25] A. T. P. Carvalho, A. F. S. Teixeira, M. J. Ramos, *J. Comput. Chem.* **2013**, 34, 1540.
- [26] O. Wise, O. Coskuner, *J. Comput. Chem.* **2014**, 35, 1278.
- [27] A. T. P. Carvalho, M. Swart, *J. Chem. Inf. Model.* **2014**, 54, 613.
- [28] M. Kaukonen, A. Gulans, P. Havu, E. Kauppinen, *J. Comput. Chem.* **2012**, 33, 652.
- [29] S. Aguilera-Segura, J. M. Seminario, *J. Phys. Chem. C* **2014**, 118, 1397.
- [30] P. Krokidas, M. Castier, S. Moncho, E. Brothers, I. G. Economou, *J. Phys. Chem. C* **2015**, 119, 27028.
- [31] S. U. Lee, J. Jung, Y. K. Han, *Chem. Phys. Lett.* **2005**, 406, 332.
- [32] S. K. Burger, M. Lacasse, T. Verstraelen, J. Drewry, P. Gunning, P. W. Ayers, *J. Chem. Theory Comput.* **2012**, 8, 554.
- [33] T. Fox., P. A. Kollman. *J. Phys. Chem. B* **1998**, 102 (41), 8070–8079.
- [34] D. Suarez, K.M. Merz. *J. Am. Chem. Soc.* **2001**, 123 (16), 3759–3770.
- [35] U. Ryde, *Proteins: Struct., Funct., Genet.* **1995**, 21 (1), 40–56.
- [36] S. C. Hoops, K. W. Anderson, K. M. Merz. *J. Am. Chem. Soc.* **1991**, 113 (22), 8262–8270.
- [37] E. J. Bautista, J. M. Seminario. Harmonic force field for glycine oligopeptides. *Int. J. Quantum Chem.* **2008**, 108 (1), 180–188.
- [38] E. B. Wilson, *J. Chem. Phys.* **1941**, 9, 76–84.
- [39] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M.

Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, G. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision B.01; *Gaussian, Inc: Wallingford, CT*, **2009**.

[40] A. D. Becke, *J. Chem. Phys.* **1993**, 98, 5648.

[41] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, 37, 785.

[42] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, 58, 1200.

[43] W. Hehre, L. Radom, P. v. R. Schleyer, J. Pople, *Ab Initio Molecular Orbital Theory*; *Wiley: New York*, **1986**.

[44] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. v. R. Schleyer, *J. Comput. Chem.* **1983**, 4, 294.

[45] M. Dolg, U. Wedig, H. Stoll, H. Preuss, *J. Chem. Phys.* **1987**, 86, 866.

[46] D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York, P. A. Kollman, AMBER 2015; *University of California: San Francisco*, **2015**.

[47] P. F. Li, K. M. Merz, *J. Chem. Theory Comput.* **2013**, 9 (6), pp 2733–2748

[48] P. F. Li, K. M. Merz, *J. Phys. Chem. B*, **2015**, 119 (3), pp 883–895

[49] A. Tafi, M. Agamennone, P. Tortorella, S. Alcaro, C. Gallina, M. Botta, *Eur. J. Med. Chem.* **2005**, 15, pp 1134–1142

List of Publications and Presentations

Publications:

Wang, R., Ozhgibesov, M. and Hirao, H.

Partial hessian fitting for determining force constant parameters in molecular mechanics.

Journal of Computational Chemistry, 2016, 37(26), pp.2349-2359.

DOI: 10.1002/jcc.24457

(Featured as the Insider Cover, DOI: 10.1002/jcc.24485)

Poster Presentations:

WANG, Ruixing; Mikhail Ozhgibesov; HIRAO, Hajime* *Partial Hessian Fitting for Quick Determination of Force Constant Parameters in Molecular Mechanics*, presented at:

1. The 9th Singapore International Chemistry Conference
11th – 14th December 2016
2. The 14th International Symposium for Chinese Organic Chemists (ISCOC) and the 11th International Symposium for Chinese Inorganic Chemists (ISCIC)
8th – 10th December 2016
3. The 4th International Conference on Asian Network for Natural & Unnatural Materials
8th – 10th June 2016