

Interfaces between Quantum Chemistry and Force Field Programs

Master Thesis

Yuan-Wei Pi



Freie Universität Berlin

Supervisor

Dr. Jan Philipp Götze

Prof. Dr. Beate Paulus

Statutory Declaration

I, Yuan-Wei Pi (皮遠激), hereby assert that my Master thesis was independently authored by myself, using the referred sources and support. I additionally assert that this thesis has not been part of another examination process. I agree that a copy of my master thesis can be borrowed from the library.

Berlin, the 10th of October, 2019

Acknowledgement

This is the end of my journey of master. I am grateful to everyone I met during my master degree. I believe who I have met, what I have learned, what I have faced become stepping stones to my next journey.

I would like to express my special thanks of gratitude to Dr. Jan Philipp Götze, who led and encouraged me to complete the thesis, and Prof. Dr. Beata Paulus, who taught me the beauty of quantum chemistry then provided me a great opportunity to do this interesting thesis. I am grateful to AG Paulus group as well, who share a friendly environment to do the research. The thankfulness is beyond description.

如今能完成碩士論文乃憑著上帝的恩典，教授與老師的諄諄教誨以及朋友和家人不吝的支持。自二〇一七年十月因緣際會從臺灣臺北來到德國柏林學習，至今已兩年。這段期間，接觸了自己沒接觸過的知識、文化和人們，了解過去管中窺豹的自己，也拓展了視野，見識了世界之大。同時，也發現自己更多可能性。

原本是學習機械工程與電機工程的我，在碩士轉而學習自然科學的基礎。這並不代表我對工程毫無熱情，反而是藉著這機會去發覺自然的美妙，探究工程背後原理的根本，進而了解萬物間的互相效力是這樣的難能可貴。這份論文，是我兩年學習與經驗的總結，有過困難與失望，也有過克服與希冀。我很感激這一路上遇見的人提供的指點與借鑒，也很感激所有貴人給予的教導與學習機會。但願這其中的知識與經驗的積累，能夠成為我的基石以及助力，幫助他人並繼續譜寫自己精彩的未來。

Abbreviations

B3-LYP	Becke 3-parameter Lee-Yang-Parr hybrid functional
BP86	Becke (B) and Perdew (P86) exchange correlation functional
C-steps	computed steps
Conjgrad	conjugate gradient
DFT	density functional theory
GHO	generalized hybrid orbital (method)
GlyAla	glycine alanine
GlySer	glycine serine
HF	Hartree-Fock (method)
LSCF	localized self-consistent field (method)
MM	molecular mechanics
NMA	normal mode analysis
PES	potential energy surface
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
RESP	restrained electrostatic potential
MP2	second order Møller-Plesset perturbation (method)
SCF	self-consistent field (method)
Steep	steepest descent
S-steps	successful steps

Contents

Abbreviations	v
List of Figures	viii
List of Tables	x
1. Introduction	1
2. Theoretical foundations	3
2.1. QM/MM method	3
2.1.1. QM and MM methods	4
2.1.2. Subtractive and additive QM/MM schemes	5
2.1.3. Interactions between the QM and MM regions	6
2.1.4. QM/MM boundary treatment	7
2.1.5. Summary	10
2.2. Geometry optimization	10
2.2.1. Steepest descent method	10
2.2.2. Conjugate gradient method	11
2.3. Normal mode analysis	11
3. Developed architecture	15
3.1. Package implementation	15
3.1.1. QM/MM calculation in the package	15
3.1.2. Implementation structure	17
3.1.3. Workflow of the package	18
3.1.4. Input files in the package	20
3.2. Testing model	24
3.2.1. Testing molecules	24
3.2.2. Partitioning in testing molecules	24
4. Results and Discussion	29
4.1. Implementation of gmx2qmmm	29
4.1.1. Package structure in gmx2qmmm	29

4.1.2. RESP problem in QM software	30
4.1.3. gmx2qmmm output files	30
4.2. Single point calculation	33
4.2.1. Energies result	33
4.2.2. Forces result	37
4.3. Geometry optimization	41
4.3.1. Steepest descent vs Conjugate gradient	41
4.3.2. Optimization with different initial step size	43
4.3.3. Optimization with different step size update factor	46
4.4. Normal mode analysis	49
5. Summary and Outlook	51
A. Appendix	53
A.1. Normal mode analysis in internal coordinates frame	53
A.2. Gram–Schmidt method	54
Bibliography	57

List of Figures

2.1. Partitioning of the QM/MM system	3
2.2. Subtractive QM/MM scheme	6
2.3. Coupling in the QM-MM boundary	8
2.4. Capping between QM-MM boundary	9
2.5. The labeling conventions in QM/MM system	9
3.1. Definition of the atoms setup in QM/MM system	16
3.2. Charge shift scheme from M_1 atom to M_2 atoms	17
3.3. Approximated model (butane model) in link scheme	18
3.4. The structure of gmx2qmmm	18
3.5. Workflow of the gmx2qmmm	19
3.6. Workflow of the optimization in gmx2qmmm	20
3.7. Workflow of the normal mode analysis in gmx2qmmm	20
3.8. Rotation illustration in GlyAla	24
3.9. Test molecules in the package	25
3.10. Visualized partitioning in GlyAla	26
3.11. Visualized point charges after charge shift scheme with different partitioning in GlyAla	26
3.12. Visualized partitioning in GlySer	27
3.13. Visualized point charges after charge shift scheme with different partitioning in GlySer	27
4.1. gmx2qmmm with classes structure	29
4.2. Graph of energy differences with different partitioning at B3-LYP level	34
4.3. GlyAla energy differences as their individual components (%) with different partitioning at B3-LYP/6-31G* level	36
4.4. GlySer energy differences as their individual components (%) with different partitioning at B3-LYP/6-31G* level	36
4.5. Error of the bond forces in the example of GlyAla at B3-LYP/6-31G* level	38
4.6. Error of the angle or dihedral forces in the example of GlyAla at B3-LYP/6- 31G* level	38

4.7. The remaining error (%) after link bond force correction in GlyAla with different partitioning at B3-LYP level	41
4.8. All computed energies for all generated geometries during a QM1-7 optimization with steepest descent and conjugate gradient	42
4.9. All computed energies for all generated geometries during a QM1-7,14-17 optimization with steepest descent and conjugate gradient	42
4.10. All computed energies for all generated geometries during a QM14-17 optimization with steepest descent and conjugate gradient	42
4.11. GlyAla geometry after geometry optimization by full-QM calculation at B3-LYP/6-31G* level	43
4.12. GlyAla geometry with different algorithms and partitioning after geometry optimization	44
4.13. Steepest descent with different initial step size	46
4.14. Conjugate gradient with different initial step size	46
4.15. Steepest descent with different step size update factors	47
4.16. Conjugate gradient with different step size update factors	47
A.1. The first two steps of the Gram–Schmidt method	55
A.2. The third step of the Gram–Schmidt method	55

List of Tables

3.1. Input files for gmx2qmmm	21
3.2. QM parameters	21
3.3. MM parameters	22
3.4. QM/MM parameters	22
3.5. Path input file	23
3.6. QM atoms with different partitioning in GlyAla	26
3.7. QM atoms with different partitioning in GlySer	27
4.1. RESP charges in Gaussian and Turbomole at BP86/STO-3G level	30
4.2. Output files in single point calculation	31
4.3. Output files during geometry optimization	32
4.4. GlyAla energy differences (after minus before rotation) with different partitioning at B3-LYP level	34
4.5. GlySer energy differences (after minus before rotation) with different partitioning at B3-LYP level	34
4.6. GlyAla energy differences as their individual components with different partitioning at B3-LYP/6-31G* level	36
4.7. GlySer energy differences as their individual components with different partitioning at B3-LYP/6-31G* level	36
4.8. GlyAla force examples at B3-LYP/6-31G* level	39
4.9. GlyAla link bond forces between C ₍₅₎ -C ₍₈₎	40
4.10. GlyAla link bond forces between C ₍₁₂₎ -C ₍₁₄₎	41
4.11. Optimization with different initial step size at B3-LYP/6-31G* level	44
4.12. Optimization with different updated step size factor at B3-LYP/6-31G* level	48
4.13. Vibrational frequencies (in cm^{-1}) for a small 18 atom test system vs a 89 atom test system	49

1. Introduction

The hybrid quantum mechanics/molecular mechanics (QM/MM) method is a state-of art technique to simulating processes in large systems, such as crystals [1] and biomolecules [2]. Balancing between accuracy and speed, it was first introduced by Arieh Warshel and Michael Levitt in 1976 [3]. Thereafter, along with Martin Karplus, they won the 2013 Nobel Prize in Chemistry for “Development of Multiscale Models for Complex Chemical Systems” [4].

Computers have been widely applied in chemistry since the 1960s [5]. At that time, the computers was not as powerful as that of processors available today; but the developers were already aware of the rising computational cost of QM calculations. They had figured out a sophisticated approach to partially overcome the cost restrictions of QM, like in the case for simulating a large biomolecular system.

QM includes the motions of atomic nuclei and electrons, to calculate details of electrons transfer and chemical processes such as bond-making or bond-breaking. We can therefore obtain the information with high accuracy but with high computational effort. On the other hand, MM simulation, using a so-called force field as the corresponding molecular potential, is described by classical mechanics. Each atom is simulated as one particle with a radius and charge. And bonded interactions are treated as springs. Consequently, compared with QM, the force field simulation is faster but lack accuracy [2,6,7].

The beauty of the QM/MM method is to maintain the advantages of both to obtain necessary information with efficiency. For the investigated system, the part where chemical reactions take place is described by the QM level, which provides higher accuracy. The remained part is treated at MM level, which is less necessary in the result. Finally, the combination of both calculations with the correction of QM/MM interaction are the result for a QM/MM calculation, which is obtained for less cost but with acceptable accuracy.

The QM/MM approach represented a significant step forward. For instance, DNA and RNA, long chain polymers, are always interesting targets for scientists to investigate the messages from genes. [8,9] The QM/MM approach made it possible to simulate those in reasonable time. M. Levitt and A. Warshel have done several researches in the reaction on DNA and RNA, such as the folding and stability of proteins and the flexibility of the furanose ring in DNA and RNA [10]. They successfully proved the QM/MM method is feasible for large biomolecular systems .

Currently, several quantum mechanics and molecular mechanics software are available.

1. Introduction

Gaussian [11], Turbomole [12] and ORCA [13] are examples for widely used QM packages. On the other hand, GROMACS [14], CHARMM [15], AMBER [16] are commonly used force field packages. Therefore, it is nowadays not difficult to run QM or MM calculations independently, and obtain required information such as energies and forces.

The goal of this thesis is to upgrade a hybrid QM/MM program, **gmx2qmmm** [17]. Thanks to these pioneers, the program can obtain calculation results from QM and MM packages, and the main task for the package is to interface between the programs. A crucial task for the **gmx2qmmm** interface is to provide an inputs for the QM and MM software and obtain the corresponding outcomes. Subsequently these outcomes are used to obtain the QM/MM result. Hence, the main challenge we have to consider is the interaction between QM and MM region.

Currently, several proposals for QM/MM schemes are available, differing often in how to treat the QM/MM boundary. For example, link atoms [18] and frozen orbitals [19] (LSCF orbitals [20] and GHO orbitals [21]) have been proposed. Among them, link atoms are a simple solution to describe the QM/MM boundary . It places a monovalent link atom such as hydrogen along the bond vector between the QM and MM atoms [7]. The link atoms scheme is therefore selected as the scheme in **gmx2qmmm**.

Python is an open source programming language which is well developed with powerful libraries [22]. For example, SciPy [23] and Numpy [24] for scientific numerical calculation, Matplotlib [25] and Pandas [26] for figure display and data visualization. It is therefore a solid choice to carry out the tasks required to build an interface between QM and MM.

Overall, **gmx2qmmm** is designed as a python script linking QM software (Gaussian, Turbomole,ORCA) and MM software (GROMACS) to perform a QM/MM calculation. It includes single point calculation, geometry optimization and normal modes analysis. Single point calculation generates the energy (and forces) of the system [27]. With the implementation of single point calculations and optimization algorithms, the geometry optimization enables us to search for a nuclear conformation with lowest energy [28]. Last but not the least, the normal mode analysis is a tool to analyze the vibrational frequencies of the system [29].

All in all, **gmx2qmmm** is designed with the feasibility to run a QM/MM calculation with acceptable accuracy in reasonable time, while making the corresponding input as trivial as possible.

2. Theoretical foundations

2.1. QM/MM method

Hybrid quantum mechanics/molecular mechanics (QM/MM) method became a popular method during the last few decades. The quantum mechanics (QM) includes electron interaction, thus the QM calculation is able to treat corresponding effects but can become time consuming. On the other hand, the molecular mechanics (MM) maintains the classical level of description, and consequently, the MM calculation is faster. It however may lack accuracy and is unable to describe processes like bond breaking [7]. The QM/MM method combines advantages of both to simulate large systems, such as enzymes [30], with acceptable accuracy and computational effort. The entire system is therefore partitioned into an inner and an outer region in the QM/MM, which is illustrated in Figure 2.1 [2]. The inner region is treated at QM level and the outer region is described by MM force fields. The inner region is usually the area where a chemical reaction takes place. The QM calculation is able to describe the changes in the electronic structure during the chemical reactions, such as bond-forming/breaking, charge transfer, electronic excitation etc [2]. The other atoms that are less important in the calculation remain in the outer region. Along with the

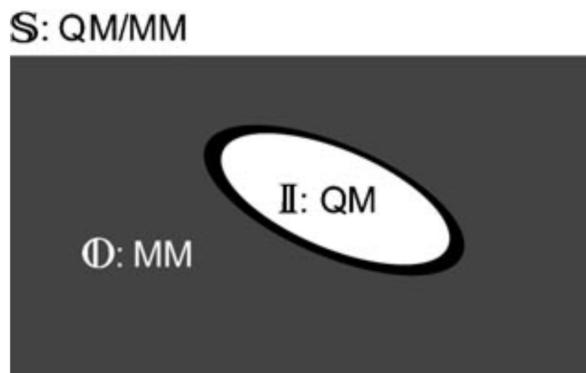


Figure 2.1.: Partitioning of the entire QM/MM system(\mathbb{S}) into inner(\mathbb{I}) and outer (\mathbb{O}) subsystems, the black ring around the inner region represents the boundary region. The figure is taken from reference [2].

QM-MM boundary region, the potential energy of the whole system includes three classes of interactions: interactions inside the QM region, interactions inside the MM region and the interactions between QM and MM regions. The interactions between two subsystems

2. Theoretical foundations

are more difficult to describe, especially if the boundary cuts through covalent bonds. Therefore, for the energy expression, several schemes have been proposed; subtractive and additive QM/MM schemes [31], which will be explained in Section 2.1.2.

2.1.1. QM and MM methods

QM method

QM methods are historically based on wave functions serving as a description for electrons, finding the solution to the Schrödinger equation with reasonable approximations [32]. The Born-Oppenheimer approximation is one important approximation assuming the nuclei remain at the same place during the calculation [33]. This is due to the fact that the mass of an atomic nucleus is much larger than that of an electron, hence the nuclei move much slower than electrons. Based on this approximation, several methods are developed, e.g. *ab initio* methods [34] and density functional theory (DFT) [35].

Ab initio methods are based on the assumption that the electrons' spatial distribution is not influenced by the movement of other electrons, which means the electron correlation is not included in the computed energy of the molecular system. In the most straightforward *ab initio* method, Hartree-Fock (HF) method, the total molecular wavefunction is approximated a Slater determinant to describe the many-body electronic wave function of the system [36]. However, the resulting energy is systematically larger than the exact solution due to the absence of electron correlation. Therefore, correlated *ab initio* methods that bring about higher accuracy have been devised, so-called post-HF methods which include the electron correlation [37].

In DFT, it is assumed that the electron distribution (density) alone can be used to obtain the energy via a density functional. Thomas and Fermi [38,39] formulated the electronic many-body system in terms of the electronic density, which enables a only three dimensional quantum mechanics system in the calculation. Thereafter, Hohenberg and Kohn elaborated this concept to show that there exists a one-to-one mapping between the electronic density and the external potential in the ground state [40]. Next, Kohn and Sham developed the Kohn-Sham method [41] that contains the exchange-correlation (XC) energy functional. Different kinds of functionals have been therefore developed, such as the Becke (B) and Perdew (P86) exchange correlation functional (BP86) and Becke 3-parameter Lee-Yang-Parr hybrid functional (B3LYP) [42].

The QM method in QM/MM should give an appropriate description of the electronic properties of the QM part, including the long-range Coulomb interactions resulting from electronic embedding (See section 2.1.3) [7]. Furthermore, due to their favorable computational efficiency, DFT methods are more widely used than *ab initio* methods in contemporary quantum chemistry and QM/MM approaches. [2].

MM method

In molecular mechanics, the potential energy include bonded terms (bond stretching, angle bending) and non-bonded terms (Lennard-Jones type van-der-Waals terms and Coulomb interaction terms) [2]. The a simple function of the potential energy can be written as

$$\begin{aligned} E_{MM} = & \sum_{bonds} k_d(d - d_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{dihedrals} k_\phi[1 + \cos(n\phi + \delta)] \\ & + \sum_{\substack{\text{non-bonded} \\ \text{pairs AB}}} \varepsilon_{AB} \left\{ \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_A q_B}{r_{AB}} \right\}. \end{aligned} \quad (2.1)$$

The symbols d, θ and ϕ are bond lengths, angles, and torsion respectively; d_0 and θ_0 are the equilibrium values for bond lengths and angles. The symbols n and ϕ are the torsional multiplicity and phase respectively; k_d , k_θ and k_ϕ are the bonded force constants. The symbols ε_{AB} and σ_{AB} are the Lennard-Jones parameters; q_A and q_B are atomic partial charges for atom A and atom B; and ε_0 is the vacuum permittivity. Currently, force fields such as AMBER, CHARMM, GROMOS and OPLS-AA are the popular examples in QM/MM [2].

2.1.2. Subtractive and additive QM/MM schemes

Subtractive QM/MM schemes

The QM/MM energy of the entire system in subtractive scheme, which is shown in Figure 2.2, is the summation of a QM calculation in QM region and an MM calculation for the entire system, and subtracting an MM calculation in QM region to avoid the double counting. The energy is therefore written as

$$E_{QM/MM}^{sub}(\mathbb{S}) = E_{MM}(\mathbb{S}) + E_{QM}(\mathbb{I} + \mathbb{L}) - E_{MM}(\mathbb{I} + \mathbb{L}). \quad (2.2)$$

The main advantage of the subtractive scheme is that the QM and MM calculation run independently; communication between two subsystems is not required. On the down side, this scheme needs complete MM parameters for the QM region, e.g., the force field needs to be flexible to describe the effect when the reaction occur. Furthermore, the electrostatic interaction is problematic owing to the absence of the polarization of the QM electron density by the MM surroundings. Therefore, the interactions between the electrons and their surroundings are typically not represented via this type of (mechanical) embedding.

Additive QM/MM schemes

To overcome these drawbacks in subtractive scheme, the requirement of flexible and complete parameters in force fields, the additive scheme is used in **gmx2qmmm**. The energy of the

2. Theoretical foundations

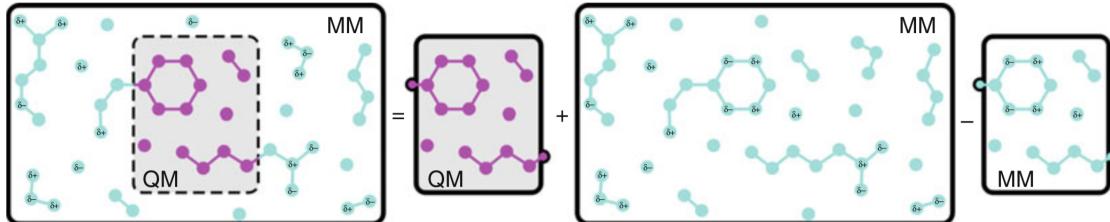


Figure 2.2.: Subtractive QM/MM scheme: The QM/MM energy is the sum of isolated QM region at the QM level and complete system at the MM level, then minus the energy of the QM region at the MM level. The figure is taken from reference [6].

additive scheme is expressed as

$$E_{QM/MM}^{add}(\mathbb{S}) = E_{MM}(\mathbb{O}) + E_{QM}(\mathbb{I} + \mathbb{L}) + E_{QM-MM}(\mathbb{I}, \mathbb{O}). \quad (2.3)$$

This shows that only the interactions among the MM region are maintained at the MM level. The interaction between QM and MM regions is specified in $E_{QM-MM}(\mathbb{I}, \mathbb{O})$, which is written as the summation of bonded, van-der-Waals, and electrostatic interaction energy between QM and MM atoms.

$$E_{QM-MM}(\mathbb{I}, \mathbb{O}) = E_{QM-MM}^b + E_{QM-MM}^{vdW} + E_{QM-MM}^{el}. \quad (2.4)$$

The details of these interaction are mentioned in next section.

2.1.3. Interactions between the QM and MM regions

The coupling of the QM and MM region is the key point to QM/MM method, especially the electrostatic interaction. To deal with the electrostatic interaction, the mechanical embedding, electrostatic embedding and polarization embedding are developed [2].

Mechanical embedding

The mechanical embedding scheme provides the interactions between the QM and MM regions at the MM level. Considering the changes of charge distribution in QM region during the reaction progresses, several accurate and various sets of MM parameters are required. Moreover, this scheme cannot represent the potential perturbation of the electronic structure of the QM region, which is resulting from the electrostatic interaction between QM and MM regions. Especially for some cases that the polarization enable to change the energetic order of the electronic excited states, this effect is crucial [7]. Therefore, the dynamical MM parameters are required in this scheme. However, updating MM parameters during the reaction progresses is unrealistic in practice. Thus, to compensate and reduce the errors, it requires larger QM region, which contradicts the original idea of the QM/MM.

Electrostatic embedding

Electrostatic embedding is a modified scheme of mechanical embedding. It contains polarization effects by including them in the one-electron terms in the QM calculation [2]. The polarization of the QM region which is caused by the MM atoms is taken into account. In the QM Hamiltonian, the charges of MM atoms are included as part of the one-electron operator:

$$H_{QM-MM}^{el} = - \sum_i^N \sum_{J \in \mathbb{O}}^L \frac{q_J}{|\mathbf{r}_i - \mathbf{R}_J|} + \sum_{\alpha \in \mathbb{I} + \mathbb{L}}^M \sum_{J \in \mathbb{O}}^L \frac{q_J Q_\alpha}{|\mathbf{R}_\alpha - \mathbf{R}_J|} \quad (2.5)$$

In this scheme, the symbols q_J are the MM point charges at position \mathbf{R}_J ; Q_α are the nuclear charges of the QM atoms at position \mathbf{R}_α ; and \mathbf{r}_i are the electron positions. The indices i , J and α run over N electrons, L point charges, and M QM nuclei, respectively [2]. Therefore, the QM-MM electrostatics interaction is treated at QM level, with only small increase in computational costs. However, there exists the risk of over-polarization around the QM-MM boundary, especially the case of the boundary crossing a covalent bond (charge shift scheme, see below). Also, the polarization in MM region from QM region is not included, which can be fixed by including (costly) polarization embedding [7].

Interaction other than electrostatic

Besides the electrostatic interaction, the interactions between QM and MM regions are bonded interactions (stretching, bending and torsional interactions), and non-bonded verder-Waals interactions, which are illustrated in Figure 2.3. These interactions are included at the MM level. Thus, calculations rely on the availability of MM parameters.

Additionally, as the discussion before, using a single set of MM parameters is one of the reasons of computational errors. But considering the computational effort, updating MM parameters for each step or enlarging the QM region are not practical solution to this issue. Hence, the suggestion for the model is still using a single set of MM parameters with additional compensation from these errors. On the other side, fortunately, the verder-Waals interactions are significant only at the short distance [2]. So it is often adequate to use a single set in the model. Overall, although it is hard to describe the system with dynamical MM parameters, a single set of MM parameters with the compensation for these errors is a practical and reasonable treatment.

2.1.4. QM/MM boundary treatment

This section is about various approaches for the case of the QM-MM boundary passing chemical bonds, such as covalent bonds. The straightforward idea to deal with this situation is the creation of one or more unpaired electrons in the QM region, which defines the system such that the boundary does not pass through a covalent bond. Hence, link atoms scheme [18] (Figure 2.4 **a**) and localized orbitals scheme [20,21] (Figure 2.4 **b,c**) are designed. To

2. Theoretical foundations

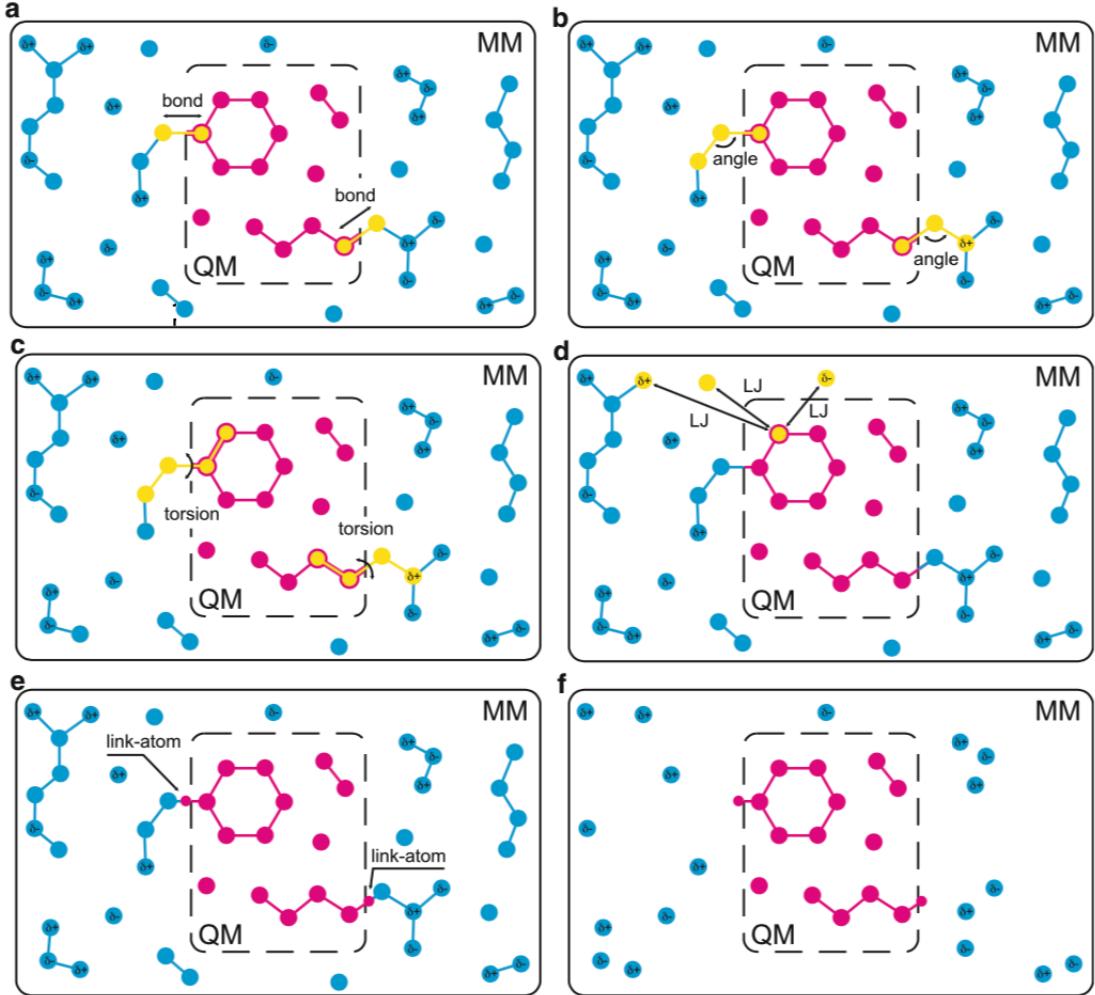


Figure 2.3.: Coupling in QM-MM boundary: Sub-figures **a – c** are the bonded interactions (bond, angle, torsion, respectively) in QM-MM boundary; Sub-figure **d** shows the van-der-Waals interactions between a QM atom and three MM atoms; Sub-figure **e** illustrates the link atoms that are placed on the Q_1 - M_1 axis; and Sub-figure **f** displays QM calculation with electrostatics embedding that link atoms method is selected as QM/MM boundary treatment, and the MM atoms that are represented point charges. The figure is taken from reference [6].

explain clearly these approaches, it is convenient to introduce the labeling conventions labeling, which is illustrated in Figure 2.5. The QM-MM boundary is located at Q_1 - M_1 . And the first shell of the QM atoms connected to Q_1 atom, are labeled as Q_2 . The QM atoms in next shell that connected to the Q_2 shell are labeled as Q_3 atoms. The same setting is in MM region.

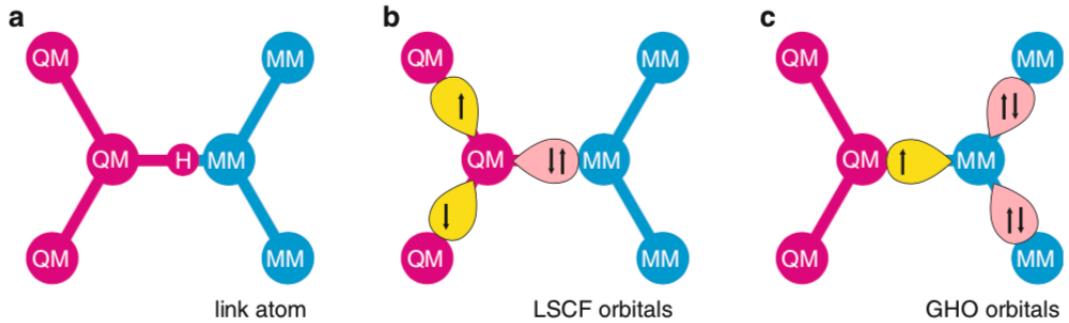


Figure 2.4.: Capping between QM-MM boundary:**a** is the link atom scheme, a hydrogen atom is selected as link atom; **b** is localized self-consistent field (LSCF) scheme; **c** is generalized hybrid orbitals (GHO). The figure is taken from reference [6].

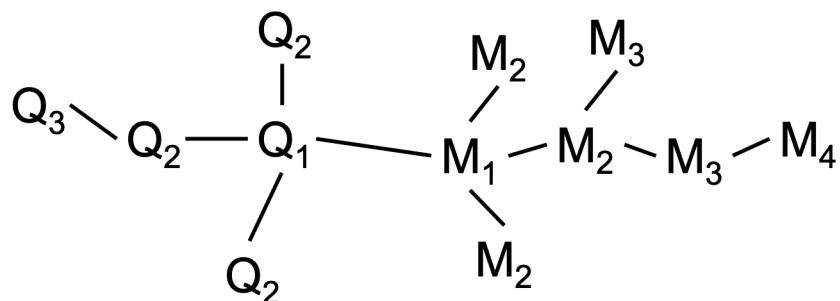


Figure 2.5.: The labeling conventions in QM/MM system, Q and M are QM atom and MM atom respectively. The QM-MM boundary is located at Q_1-M_1 .

Link atoms

The easiest way to treat the covalent bonds cut by the QM-MM boundary is to introduce a monovalent link atom, usually a hydrogen atom [2], which is shown in Figure 2.4.a. It is not part of the real system. The link atoms are present only during the QM calculation, not in the MM region. To avoid extra degrees of freedom of the link atom, the location of it is along the Q_1-L axis. And the Q_1-L distance R_{Q_1-L} is scaled with respect to Q_1-M_1 distance $R_{Q_1-M_1}$ by a scaling factor c :

$$R_{Q_1-L} = cR_{Q_1-M_1}. \quad (2.6)$$

The scaling factor can be determined with geometry optimization of model compounds [7]. In the QM geometry optimization, the link atoms should be force-free. In short, the scaling factor is obtained from the converged, optimized geometries.

Another problem that has to be solved in link atom scheme is the over-polarization of the Q_1-L bond due to the M_1 point charge nearby. The simplest way to resolve this problem is to ignore the M_1 charge by setting it to zero. But the total charges and the dipole field

2. Theoretical foundations

of system are required to maintain the same as in the original system, thus the charge shift scheme has been developed. The M_1 charge is partitioned equally and redistributed to the M_2 atoms. Thereafter, an additional pair of opposite point charges and distance adjusted in order to compensate for the difference in dipole field caused from the initial shift.

2.1.5. Summary

Aforementioned, obtaining dynamical parameters in MM force fields is complicated and still in progress. Therefore, the better model for the QM/MM calculation is the electrostatics embedding with link atoms in the QM-MM boundary. It requires only a single set of MM parameters with the care of the compensation the errors in link atoms scheme to provide reasonable accuracy and effort.

2.2. Geometry optimization

Geometry optimization is the process to generate the optimal structure, the equilibrium geometry, of a system with energy minimization on the potential energy surface (PES). Due to the lack of the information of the PES, the constraints of the energy function is unclear beforehand. That is to say, geometry optimization is an unconstrained optimization problems that minimize the energy functions with no restrictions on real variables [28]. The intuitive idea for this optimization is to iterate over the different conformation numerically until the energy of the molecule has reached the local minimum that is closest to the starting point. The steepest descent and conjugate gradient methods are widely selected as algorithms in geometry optimization [11].

2.2.1. Steepest descent method

Steepest descent, which is also called gradient descent, is an algorithm for finding the nearest local minimum of a function, which searching direction is along with the negative gradient. The algorithm searches for a stationary point of the PES, with the direction of forces, which are the negative gradient of the PES [28]. The length of the searching direction, which is so-called step size, is adjusted during the iteration. The criteria of the update step size is followed by

$$\text{if } E_{n+1} < E_n, \quad h_{n+1} = (1 + \alpha)h_n \quad (2.7)$$

$$\text{if } E_{n+1} \geq E_n, \quad h_n = \alpha h_n \quad (2.8)$$

$$0 \leq \alpha \leq 1$$

where the symbol α is the update step size factor. If the current energy is smaller than the previous one, the optimizer tends to increase the step size. Conversely, if the current energy is larger or equal to the previous one, which means a potentially correct searching direction, the step size then becomes smaller to search for the near position.

We describe all $3N$ coordinates as the vector \mathbf{r} with the initial displacement h_0 of the

gradient. Therefore, coordinates are updated as

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \frac{\mathbf{F}_n}{\max(|\mathbf{F}_n|)} h_n \quad (2.9)$$

where the symbol h_n , the step size, is the maximum displacement and the \mathbf{F}_n is the force at step n . And the $\max(|\mathbf{F}_n|)$ is the largest magnitude of the force component at step n [11].

2.2.2. Conjugate gradient method

Similar to the steepest descent, the conjugate gradient updates coordinates with negative gradient of the function and the adjusted step size. In addition, it contains the previous searching direction, the negative gradient of the function used for the previous step update. This algorithm is suitable for large sparse systems and can be used to solve unconstrained optimizations [28]. The Fletcher-Reeves conjugate gradient method [43] is frequently used for this case. The coordinate in geometry optimization with Fletcher-Reeves conjugate gradient method is updated as

$$\mathbf{r}_{n+1} = \mathbf{r}_n - \mathbf{k}_n h_n \quad (2.10)$$

$$\mathbf{k}_{n+1} = -\mathbf{F}_n + \gamma_n \mathbf{k}_n \quad (2.11)$$

$$\gamma_n = \frac{|-\mathbf{F}_n|^2}{|-\mathbf{F}_{n-1}|^2} \quad (2.12)$$

The symbol h_n is the step size at step n ; the vector \mathbf{k}_n and the vector \mathbf{F}_n are the searching direction and the force at step n , respectively; γ_n is the factor for the previous searching direction at step n [44].

2.3. Normal mode analysis

Normal mode analysis (NMA) is used to describe the fundamental frequencies and motions of an oscillating system. The internal motions of a molecule can be described as the superposition of these individual movements, which are also called normal modes [45,46].

Without external field, the energy of the molecule does not depend on its translational and rotational degrees of freedom. The potential energy therefore depends on vibrational degrees of freedom, $3N - 6$ (or $3N - 5$ for linear molecules) [29]. The potential energy, V is the function of a set of coordinates q_i , i.e., $V = V(q_1, q_2, q_3, \dots, q_n)$. With the Taylor expansion, we can expand V in terms of q_i ,

$$V = V(0, 0, 0, \dots, 0) + \sum \left(\frac{\partial V}{\partial q_i} \right)_0 q_i + \frac{1}{2} \sum \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 q_i q_j + \dots \quad (2.13)$$

Since the geometry is typically located at a stationary point of the PES before running a NMA, the first derivatives (formula) can be set to zero. Neglecting the actual energy of the

2. Theoretical foundations

system (only relative energies), the potential energy in harmonic approximation becomes

$$V = \frac{1}{2} \sum \left(\frac{\partial V}{\partial q_i \partial q_j} \right)_0 q_i q_j \quad (2.14)$$

In short, the calculations require the second derivatives of the potential, which is a Hessian matrix. The Hessian matrix is obtained from the second derivatives of the potential with respect to coordinates \mathbf{q} , which is defined as

$$H_{ij} = \left(\frac{\partial V}{\partial q_i \partial q_j} \right)_0 \quad (2.15)$$

Consequently, the equation of motion can be expressed as

$$\mathbf{M} \frac{d^2 \Delta \mathbf{q}}{dt^2} + \mathbf{H} \Delta \mathbf{q} = 0 \quad (2.16)$$

where the diagonal matrix \mathbf{M} includes the masses of particles. Because of the three Cartesian coordinates for each particle, the mass is repeated three times in the matrix. A solution to the Eq.2.16 is the $3N$ -dimensional vector

$$\mathbf{u}_k(t) = \mathbf{a}_k e^{-i\omega_k t} \quad (2.17)$$

where \mathbf{a}_k is a complex vector of amplitude and phase information, and the ω_k is the frequency that the motion is represented by \mathbf{u}_k . Therefore, the equation of motion in Eq.2.16 can be written as an eigenvalue equation

$$\mathbf{H} \mathbf{u}_k = \omega_k^2 \mathbf{M} \mathbf{u}_k \quad (2.18)$$

The complete set of the solution \mathbf{u}_k , $1 \leq k \leq 3N$, can be expressed as the respective columns of the matrix \mathbf{U} . And the set of elements $\lambda_k = \omega_k^2$ can be represented as a diagonal matrix Λ . Thus, the Eq.2.18 can be expressed as

$$\mathbf{H} \mathbf{U} = \mathbf{M} \mathbf{U} \Lambda \quad (2.19)$$

Moreover, to transform to a standard eigenvalue equation expression, $\widetilde{\mathbf{H}} \widetilde{\mathbf{U}} = \widetilde{\mathbf{U}} \Lambda$, we introduce mass weighted coordinates,

$$\widetilde{\mathbf{U}} = \mathbf{M}^{\frac{1}{2}} \mathbf{U} \quad (2.20)$$

$$\widetilde{\mathbf{H}} = \mathbf{M}^{-\frac{1}{2}} \mathbf{H} \mathbf{M}^{-\frac{1}{2}} \quad (2.21)$$

The symbol $\widetilde{\mathbf{H}}$ is so-called mass-weighted Hessian matrix; $\widetilde{\mathbf{U}}$ is an orthonormal basis set, which means $\widetilde{\mathbf{U}}^T \widetilde{\mathbf{U}} = 1$. These are the normal modes of the system [29].

But to obtain the normal mode frequencies, the choice of the internal coordinates and the elimination of the translating the rotating coordinates are required. Although the Cartesian coordinates are universal and least ambiguous, the x,y and z coordinates lack description of the chemical structure and bonding of a molecule [28]. Therefore, the internal coordinates are introduced to describe the molecular structure with the advantage of

2.3. Normal mode analysis

removing the translating the rotating coordinates. The dimension of the matrix is therefore reduced to $3N - 6$ (or $3N - 5$ for a linear system) modes for vibrational analysis [11]. The details of transformation into internal coordinates are described in Appendix A.1. The idea is to diagonalize an internal coordinates Hessian matrix \mathbf{H}_{INT} which then yields the vibrational wavenumbers from its eigenvalues.

3. Developed architecture

3.1. Package implementation

The package **gmx2qmmm** is designed to run a QM/MM calculation with acceptable accuracy in reasonable time. The illustration below is the updated implementation of **gmx2qmmm 1.0** [17]. At the same time, input conventions are kept close to those of the GROMACS package, allowing for an easy transition for the unexperienced user.

3.1.1. QM/MM calculation in the package

gmx2qmmm uses the additive scheme for energy and forces calculation with link atoms approach in case of covalent bonds breaking. The system setup and labeling of atoms are shown in Figure 3.1. Atoms in the system are either QM atoms (notation is Q in figure) or MM atoms (notation is M in figure). If link atoms exist, they are connected to the Q_1 atom and substituted for the M_1 atom during the QM calculation. The total energy and forces are calculated in Eq.3.1 and Eq.3.2.

$$E_{QM/MM} = E_{QM,polar} + E_{MM,mod} - E_{link} \quad (3.1)$$

$$F_{QM/MM} = F_{QM,polar} + F_{MM,mod} - F_{link} \quad (3.2)$$

QM calculation

In the QM calculation, MM atoms are represented with their point charges and the M_1 atoms are replaced with link atoms. The distance between L and Q_1 is approximated as

$$R_{L-Q1} = c \cdot R_{M1-Q1} \quad (3.3)$$

$$c = 0.71290814$$

where c is the ratio between optimized butane central $C - C$ bond length (R_{M1-Q1}) and the corresponding optimized $C - H$ bond length (R_{L-Q1}) at B3-LYP/6-31G* level, where the H atom is the link atom that replaced the original C atom. The butane model is illustrated in Figure 3.3 and will be explained later. Moreover, to maintain overall charge of the system and the dipole field of the M atoms during the QM calculation, the difference between total charge of atoms and the QM atoms is equally distributed over all M_1 atoms. After the substitution of M_1 atom, the charge shift scheme is applied to the MM region; the charge of M_1 atom is evenly distributed to the M_2 atom. This process is illustrated

3. Developed architecture

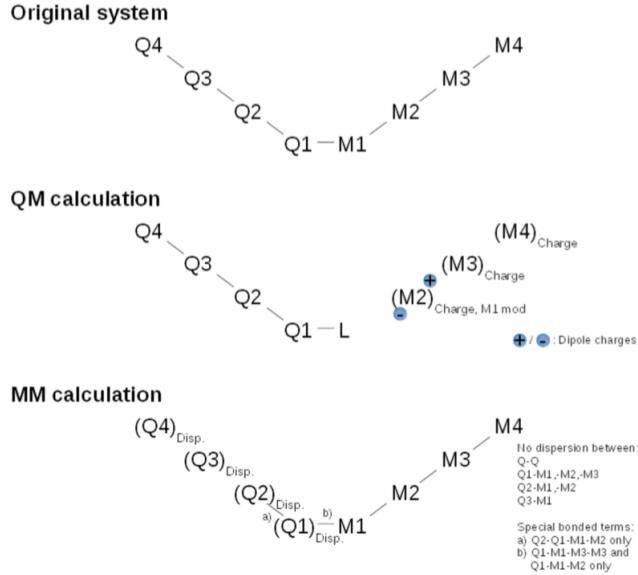


Figure 3.1.: Definition of the atoms setup in QM/MM system; the orginal system can be partitioned into QM and MM region then run the QM and MM calculation, respectively. The figure is taken from reference [17].

in Figure 3.2. However, these series of moves have changed the dipole moment on each Q atom. Therefore, two point charges are placed along the M_1 - M_2 axis with opposite charge. The location of these charges are obtained by the numerical optimization that adjust the overall dipole field as close as possible to the original dipole field.

While the charge of these opposite charge is set to an arbitrary value of $+/-$ (Figure 3.1), the dipole field is numerically optimized to mimic the original MM dipole situation.

The geometry of Q atoms and L atoms (if present) with the charge-shifted point charge field is the input of the QM calculation. The output of the QM calculation provides the $E_{QM,polar}$, $F_{QM,polar}$ and the atomic charge of the Q atoms for QM/MM calculation afterwards. The use of the atomic charge of Q atoms is in link energies and forces calculation, which will be explained later.

MM calculation

In MM calculation includes all atoms in the system. In order to prevent double counting, the Coulomb interaction is removed between Q - Q and Q - M , QM atom charge is set to 0. Also, the dispersion interactions are removed between Q - Q , Q_1 - M_2 , Q_1 - M_3 , Q_2 - M_2 . Moreover, due to the L atom correction, all Q - M_1 dispersion interactions does not exist in the MM calculation. For bonded terms involving Q atoms, the MM energy contains only Q - M - M , Q - Q - M - M and Q - M - M - M bonded terms for respective force field parameters. With these settings, the output of the MM calculation are $E_{MM,mod}$ and $F_{MM,mod}$.

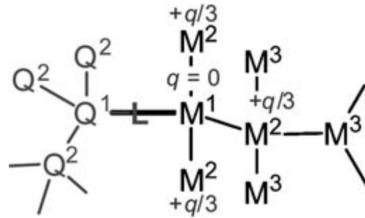


Figure 3.2.: Charge shift scheme from M_1 atom to M_2 atoms; the charges are equally distributed to M_2 atoms. The figure is taken from reference [2].

Link atom energy and forces

If the partitioning does not cut any covalent bonds, meaning absence of L atoms. The total energy and forces in Eq. 3.1 and 3.2 do not contain the E_{link} and F_{link} term, respectively. On the other hand, if L atoms exist, a correction for the total energy and forces is required. Therefore the link scheme is applied as a approximation with a butane ($CH_3CH_2CH_2CH_3$) cutting model. The division took place at the central $C - C$ bond. Next the central carbon atom is replaced with a hydrogen atom. The cutting position and replacement model is shown in Figure 3.3. With the approximated butane model, the corresponding energy for the correction is written as

$$E_{corr}(r) = -0.06516422r^3 + 0.62589230r^2 - 1.98303692r - 76.55202164 \quad (3.4)$$

Forces are a property applying to the individual atom, not the total system as for the energy. Thus, the forces are corrected separately as Q_1 atom in Eq. 3.5 and L atom in Eq.3.6.

$$F_{Q_1corr}(r) = -0.12165160r^3 + 1.14556935r^2 - 3.56253304r + 3.65070551 \quad (3.5)$$

$$F_{Lcorr}(r) = +0.12734680r^3 - 1.21401610r^2 + 3.84474061r - 4.04529919 \quad (3.6)$$

The energy and forces approximation function are 3^{rd} order fitting functions of the distance dependent difference between a $C-H$ and $C-C$ stretch in butane model system, with or without cut, respectively.

3.1.2. Implementation structure

gmx2qmmm is written in python 2.7, the structure of the package is displayed in Figure 3.4. To execute the script, the access to QM software and MM software is required. The QM software includes Gaussian16, Turmole 7.3, ORCA 4.1.2, on the other hand, the MM software contains Gromacs 2019. Python translates the input into required input for these software. The QM and MM software run their tasks separately. After the QM and MM calculation, the package reads the output and performs the QM/MM calculation, generating the energies and forces.

3. Developed architecture

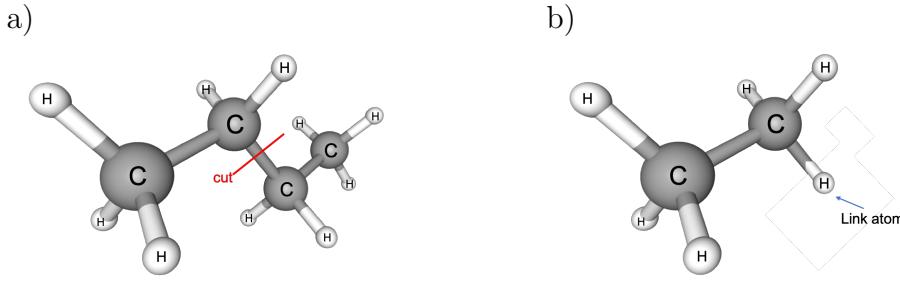


Figure 3.3.: Approximated model (butane model) in link atom scheme: a) Cutting between the butane C-C bond, b) The substitution C atom by H atom as link atom after cutting

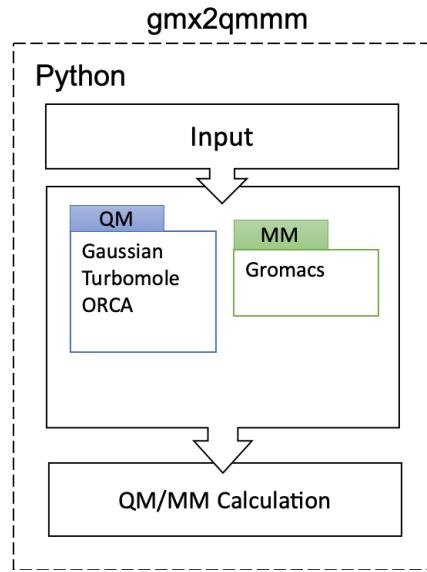


Figure 3.4.: The structure of **gmx2qmmm**

The available jobs in **gmx2qmmm** are single point calculation, geometry optimization and normal mode analysis. The processes and workflow of these jobs will be mentioned in next subsection.

3.1.3. Workflow of the package

The workflow of the **gmx2qmmm** is illustrated in Figure 3.5. After the initialization, single point calculation, geometry optimization and normal mode analysis are available.

The initialization, which is showed in Figure 3.6, requires the coordinates of the system, the partitioning of the QM and MM regions, the preparation of the charge shift, link atoms schemes as inputs. Also, the detailed parameters for QM and MM calculation are also needed. Afterwards, the program will always at least run a single point calculation.

With the data from a single point calculation, geometry optimization is possible given specific criteria for the iteration cycles. In the optimization, each iteration performs a

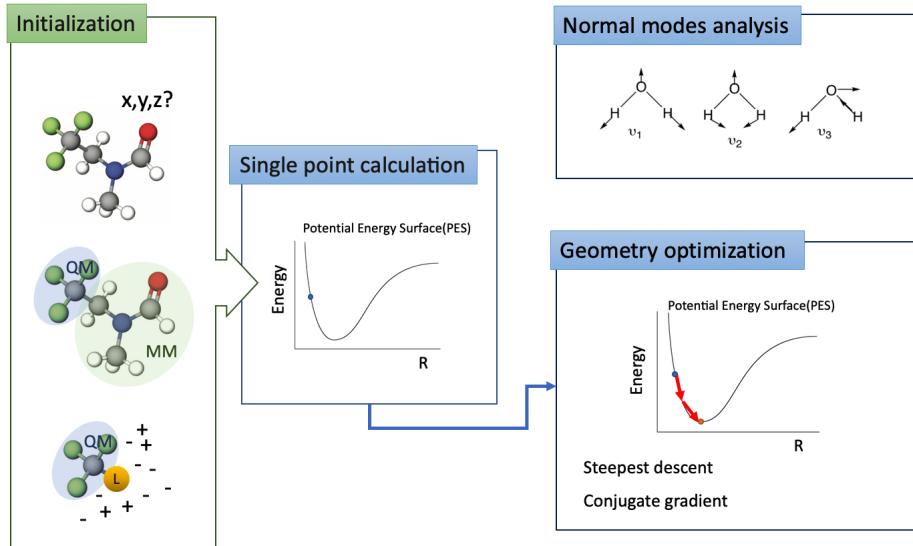


Figure 3.5.: Workflow of the `gmx2qmmm`

single point calculation to get the current energy and forces of all atoms. The propagator, steepest descent or conjugate gradient, then creates a new geometry for next iteration if the maximum of the force is larger than the force threshold or the number of iterations does not reach the number of maximum cycles. The new geometry is updated based on the algorithms mentioned in Section 2.2. The default step size update factor during the optimization is referred to the value in GROMACS optimization. Consequently, the propagator, maximum number of iterations, initial step size and force threshold are changeable parameters for the optimization. The workflow of the geometry optimization is also illustrated in Figure 3.6.

The purpose of normal mode analysis (NMA) is to analyze uncoupled molecular vibrations and their associated fundamental frequencies. The output is obtained from the mass weighted Hessian matrix. The computational processes are followed by the equations in Section 2.3 and illustrated in the Figure 3.7.

3. Developed architecture

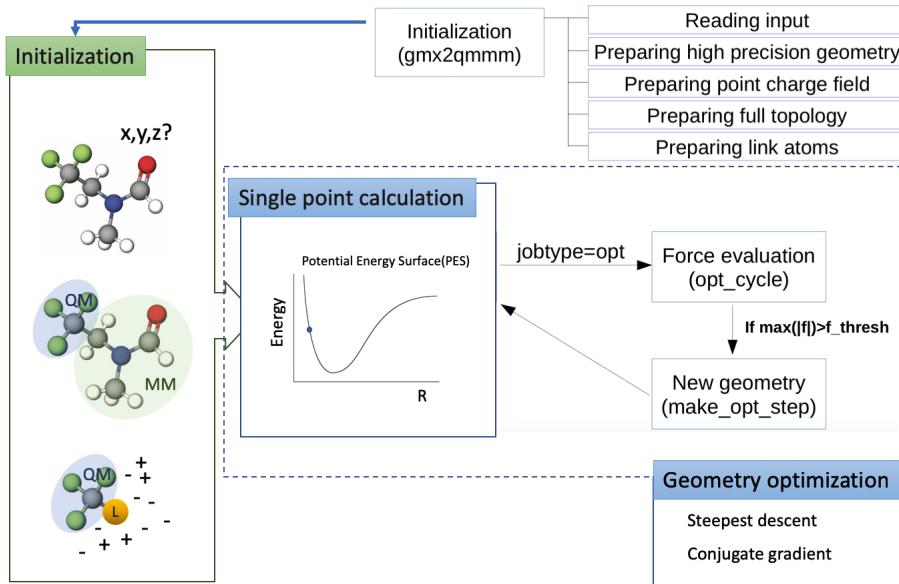


Figure 3.6.: Workflow of the optimization in **gmx2qmmm**

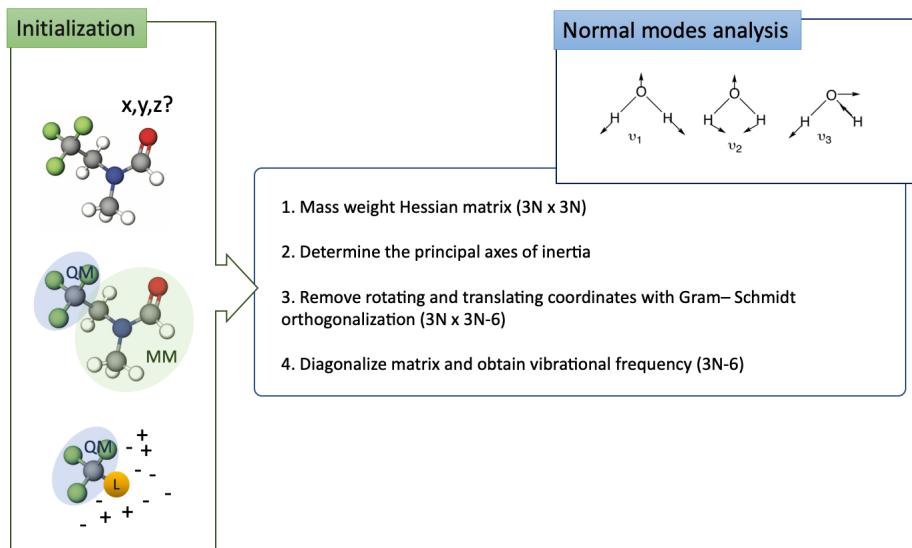


Figure 3.7.: Workflow of the normal mode analysis in **gmx2qmmm**

3.1.4. Input files in the package

The required input files in the package are listed in Table 3.1. If the parameters are not entered in the input files, the program will run the job with the default values.

Table 3.1.: Input files for gmx2qmmm

Input files	Command	Default input name
a) Molecular structure file(.pdb)	-pdb	mol.pdb
b) QM atoms file(.ndx)	-n	qmatoms
c) QM parameters (.dat)	-qm	qm
d) MM parameters (.dat)	-mm	mm
e) QM/MM parameters (.dat)	-qmmm	qmmm
f) Active atoms (.ndx)	-act	active_atoms
g) Path file (.dat)	-path	path
h) Logfile (.log)	-g	logfile

The executed example with default parameters is

```
python ~/gmx2qmmm/gmx2qmmm.py -pdb mol.pdb -n qmatoms -qm qm -mm
mm -qmmm qmmm -act active_atoms -path path -g logfile
```

a) Molecular structure file

To generate the conformation and topology files for the program, a *.pdb* file is required. **gmx2qmmm** runs **gmx pdb2gmx** to generate *.gro* and *.top* file.

b) QM atoms file

The QM atoms file contains the list of numbers in any order, indicating which atom is a QM atom during the calculation.

c) QM parameters

QM parameters are the setting for QM calculation, including selected program, method, basis, charge and multiplicity of QM subsystem, cores and memory for computational environment. The details are shown in Table 3.2.

Table 3.2.: QM parameters

Parameters with default value	Description
program = G16	Selected QM program
method = BP86	Selected QM method; any choice that allows for point charge gradients is valid
basis = STO-3G	Selected basis set in QM

3. Developed architecture

charge = 0	Total charge of the QM region after removal of the environment
multiplicity = 1	Multiplicity of the QM region after removal of the environment
cores = 1	Number of computing cores during QM calculation, >1 implies SMP parallel job
memory = 1000	Total memory for the QM calculation, in MB

d) MM parameters

MM parameters are the setting for MM calculation, including the van-der-Waals radius and the force field. Users enable to import their modified force field for the calculation with the path of the force directory, which the suffix of the directory is *.ff*.

Table 3.3.: MM parameters

Parameters with default value	Description
rwdw = 2.0	van-der-Waal radius, in nm
force_field = AMBER99SB-ILDN	Selected force field in MM calculation. Users can also import their modified force field for the calculation with the path of the force field directory (<i>.ff</i>).

e) QM/MM parameters

QM/MM parameters are the setting of the calculation, including the job name and the job type. If the program runs an optimization calculation (jobtype=opt), users are able to enter the maximum cycle of iterations, initial step size, and force threshold for the optimized criteria. The details are shown in Table 3.4.

Table 3.4.: QM/MM parameters

Parameters with default value	Description
jobname = testjob	Name of the job

jobtype = singlepoint	Jobtype of the calculation. “singlepoint” and “opt” are available, which stand for single point calculation and geometry optimization.
propagator = steep	Read if the jotype = opt, the methods “steep” (steepest descent) and “conjgrad”(conjugate gradient) are available.
maxcycle = 5	Maximum number of optimization iterations, which satisfies the iteration criteria
initstep = 0.1	Initial stepsize for the optimization , in bohradii
f_thresh = 0.00001	Force threshold for the optimization, in hartree/bohradius

f) Active atoms

This file contains active atoms in the calculation. The format is same as b). The atoms in this file are “active” in the calculation. The forces of inactive atoms are set to zero before they are used for any purpose.

g) Path file

The path file includes the path of the specific QM and MM software and the executed command for the certain software. Input paths become global variables for the whole package run. It is possible for users to export the software from shell. If the user loads the software via the shell, the path input file is not needed. But in this case the command for the system is still needed if the command differs from the default. The default path file is an empty file. The details are shown in Table 3.5.

Table 3.5.: Path input file

Parameters with default value	Description
g16path	The path of the Gaussian16
g16cmd = “rung16”	The executed command in Gaussian16
tmpath	The path of the Turbomole
tmcmd = “rung16”	The executed command in Turbomole
orcopath	The path of the ORCA
orcacmd = “rung16”	The executed command in ORCA

3. Developed architecture

gmxpath	The path of the GROMACS
gmxcmd =“gmx19”	The executed command in GROMACS

h) Logfile

The logfile displays every action of **gmx2qmmm** with a time stamp. Users are able to find here all errors that are not related to the python code, such as wrong settings of failed calculations. For example, the wrong setting or failed calculation in either the QM calculation or the MM calculation.

3.2. Testing model

3.2.1. Testing molecules

The calculated systems are Glycine Alanine (GlyAla), Glycine Serine (GlySer). In order to investigate the program result from the energy aspect, the rotation of the methyl group ($-CH_3$) in GlyAla and the rotation of the hydroxymethyl group ($-CH_2OH$) in GlySer are also included. Both groups are rotated 60° clockwise without changing $C - H$ bond lengths. The rotation illustration took GlyAla as an example shown in Figure 3.8. Besides the rotating group, the geometry is remained the same coordinates as before; bond lengths, angles and dihedral angles maintain the same. The geometries of all calculated systems are illustrated in Figure 3.9.

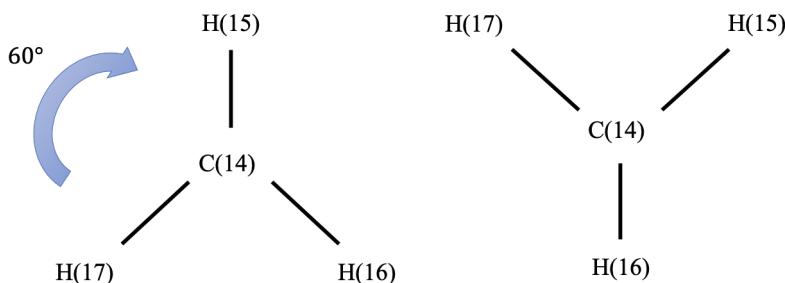


Figure 3.8.: Rotation illustration in GlyAla, the methyl group is rotated 60° clockwise without changing $C - H$ bond lengths.

3.2.2. Partitioning in testing molecules

Each molecule is subject to three types of partitioning. The first partitioning in both molecules only includes $-CH_2NH_2$ in the QM region. The cut occurs at $C_{(5)}-C_{(8)}$. The $C_{(8)}$ (M_1 atom) is replaced by H atom (L atom) in QM calculation. The second partitioning contains extra methyl group ($-CH_3$) in GlyAla, extra hydroxymethyl group ($-CH_2OH$)

3.2. Testing model

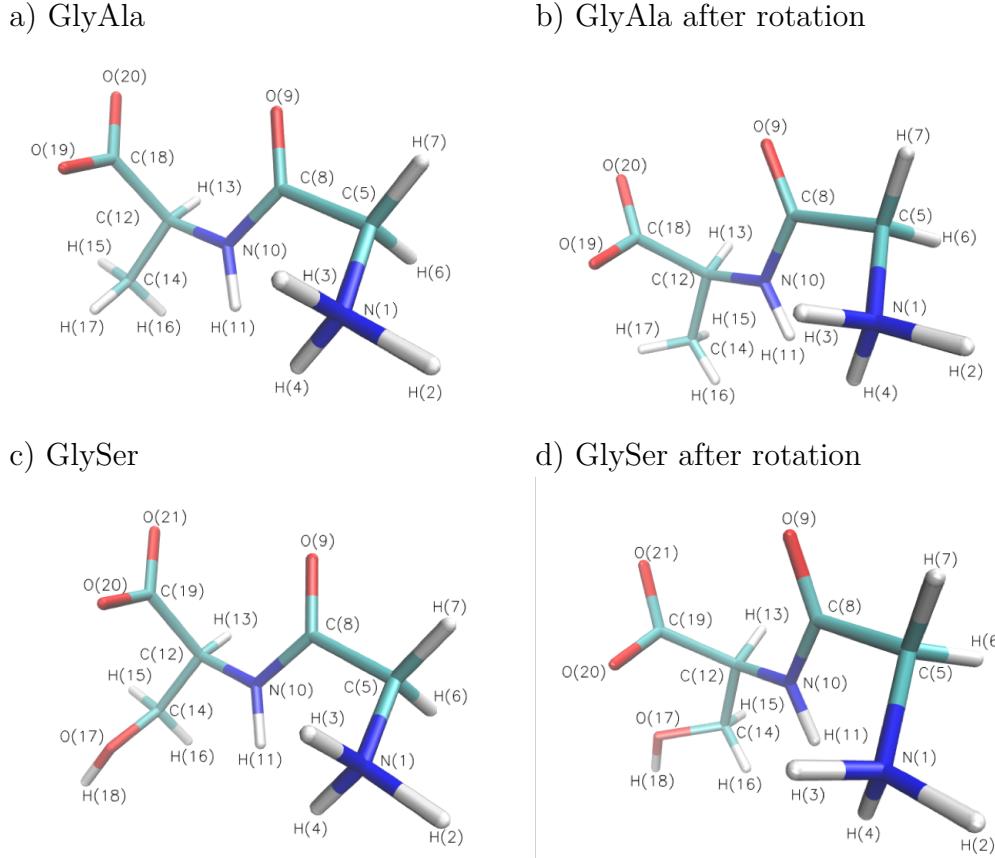


Figure 3.9.: Testing molecules in the package

in GlySer. The covalent bonds $C_{(5)}-C_{(8)}$ and $C_{(12)}-C_{(14)}$ cross the QM/MM boundary. Therefore, there are two link atoms in these calculations. The third partitioning only contains methyl group as QM atoms in GlyAla, and hydroxymethyl group as QM atoms in GlySer. The link atom substitution only takes place at the $C_{(12)}-C_{(14)}$ bond. The details of the partitioning are shown in Figure 3.6 and 3.7, as well as the illustration in Figure 3.10 and 3.12.

The Figures 3.11 and 3.13 illustrate the point charges after the charge shift operations in the QM calculation with different partitioning. The processes of charge shifting are mentioned in Section 3.1.1. In the QM calculation, in addition to M_1 atom, MM atoms are represented as point charges. Also, the extra pair of point charges are near the M_2 atoms for the purpose of the compensation for modification of the dipole field from charge shifting. Note that because of two QM regions in the partitioning 2, there exist two pairs of point charges around M_2 atom, which is a nitrogen ($N_{(10)}$) in our case.

3. Developed architecture

Table 3.6.: QM atoms with different Partitioning in GlyAla

Partitioning	QM atoms
1 (QM1-7)	N ₍₁₎ ,H ₍₂₎ ,H ₍₃₎ ,H ₍₄₎ ,C ₍₅₎ ,H ₍₆₎ ,H ₍₇₎
2 (QM1-7,14-17)	N ₍₁₎ ,H ₍₂₎ ,H ₍₃₎ ,H ₍₄₎ ,C ₍₅₎ ,H ₍₆₎ ,H ₍₇₎ , C ₍₁₄₎ ,H ₍₁₅₎ ,H ₍₁₆₎ ,H ₍₁₇₎
3 (QM14-17)	C ₍₁₄₎ ,H ₍₁₅₎ ,H ₍₁₆₎ ,H ₍₁₇₎

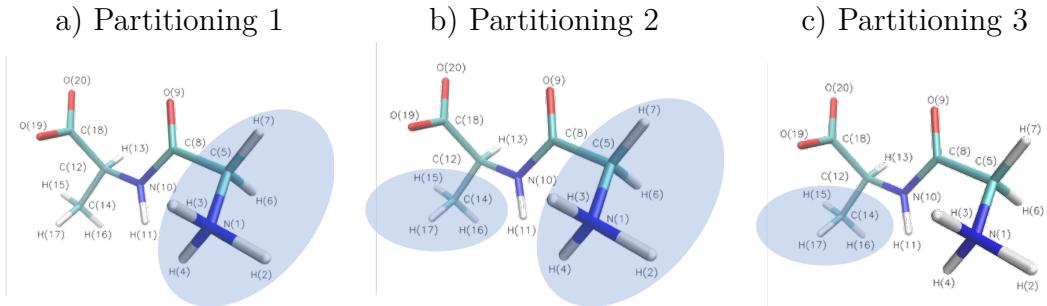


Figure 3.10.: Visualized partitioning in GlyAla, the blue part stands for QM region. a) GlyAla Partitioning 1, b) GlyAla Partitioning 2, c) Partitioning 3.

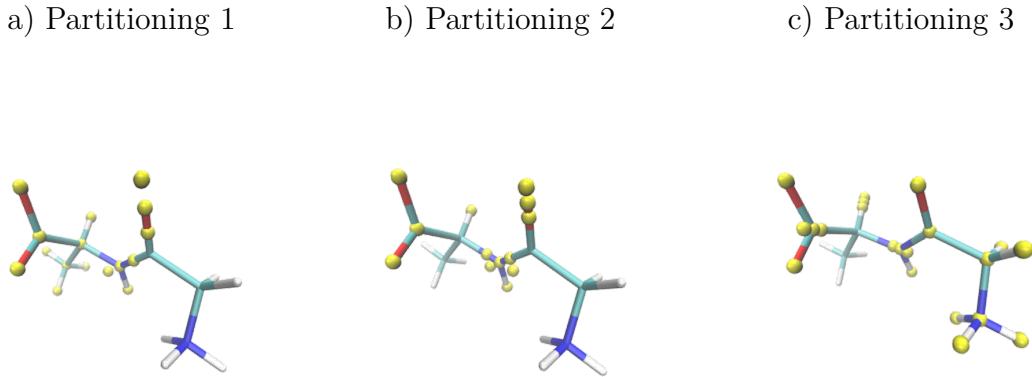
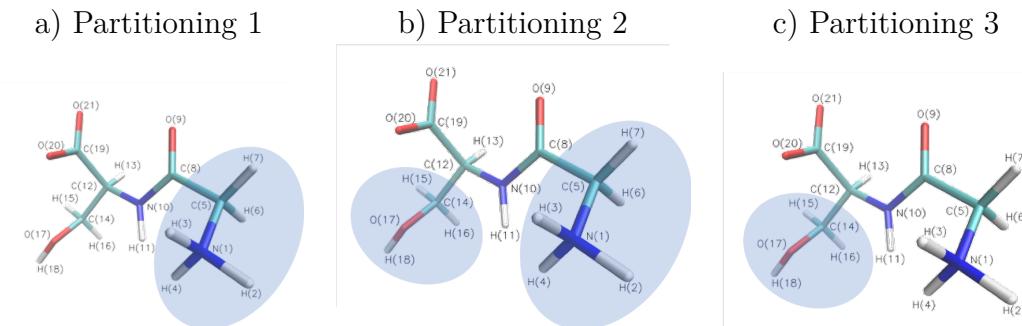
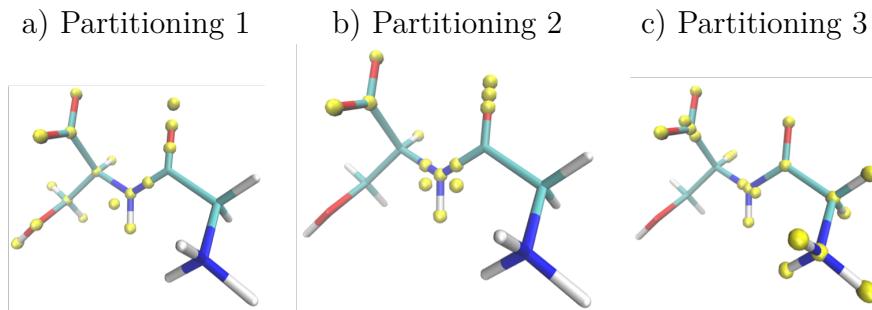


Figure 3.11.: Visualized point charges after charge shift scheme with different partitioning in GlyAla, the yellow points stand for point charge. a) GlyAla partitioning 1, b) GlyAla partitioning 2; the pair atoms on H₍₁₃₎ and C₍₁₈₎ are too close to their parent atoms to be distinguished visually , c) GlyAla partitioning 3.

Table 3.7.: QM atoms with different Partitioning in GlySer

Partitioning	QM atoms
1	N ₍₁₎ ,H ₍₂₎ ,H ₍₃₎ ,H ₍₄₎ ,C ₍₅₎ ,H ₍₆₎ ,H ₍₇₎
2	N ₍₁₎ ,H ₍₂₎ ,H ₍₃₎ ,H ₍₄₎ ,C ₍₅₎ ,H ₍₆₎ ,H ₍₇₎ , C ₍₁₄₎ ,H ₍₁₅₎ ,H ₍₁₆₎ ,O ₍₁₇₎ ,H ₍₁₈₎
3	C ₍₁₄₎ ,H ₍₁₅₎ ,H ₍₁₆₎ ,O ₍₁₇₎ ,H ₍₁₈₎


Figure 3.12.: Visualized partitioning in GlySer, the blue part stands for QM region. a) GlySer Partitioning 1, b) GlySer Partitioning 2, c) Partitioning 3.

Figure 3.13.: Visualized point charges after charge shift scheme with different partitioning in GlySer, the yellow points stand for point. a) GlySer Partitioning 1, b) GlySer Partitioning 2; the pair atoms on H₍₁₃₎ and C₍₁₉₎ are too close to their parent atoms to be distinguished visually, c) Partitioning 3.

4. Results and Discussion

This chapter presents the implementation result of **gmx2qmmm** and the testing result for the single point calculation, geometry optimization and normal mode analysis.

4.1. Implementation of gmx2qmmm

4.1.1. Package structure in gmx2qmmm

Currently, the **gmx2qmmm** runs the entire calculation by calling functions in sequence. Therefore, the input variables are carried from each function to the next, even if the current function does not require all available variables. Such an approach has the risk of using the wrong variable due to mixing up the names. Also it is difficult to obtain information from a running process. A better approach to organize the data is the utilization of classes. Input files are organized as a class with corresponding functions. Some set variable can become a local variable for a whole class. A class (Info) can be defined that contains the information for each step. The calculations in QM, MM and QM/MM are also written as classes such that the main class has access to the functions in Class QM, MM and QMMM. Consequently, the information becomes accessible and reproducible, which is especially useful when trying to obtain information from previous optimization steps. Moreover, extending the program by additional functions is easier with this structure in the future. The general idea of the classes structure is illustrated in Figure 4.1.

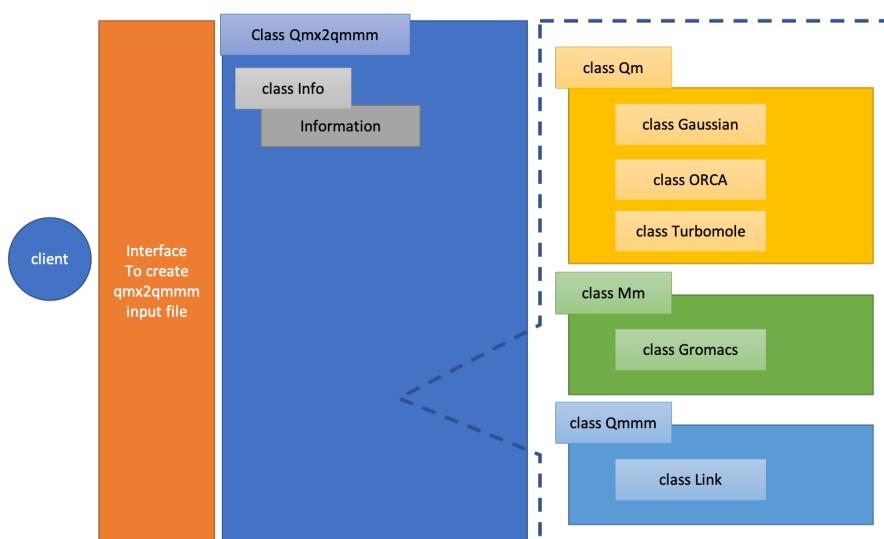


Figure 4.1.: gmx2qmmm with classes structure

4. Results and Discussion

4.1.2. RESP problem in QM software

The link scheme calculation involves atomic charges in QM region, which are obtained from RESP (Merz Kollman) scheme [47,48]. The RESP approximation is available in Gaussian and Turbomole, but not in ORCA [13]. Moreover, RESP values from Turbomole are not comparable to those obtained from Gaussian. The testing model in Table 4.1 is 8 atoms from GlyAla in QM calculation (Partitioning 1). The total charge of these 8 atoms should be equal to 1.00000. However, the result from Turbomole is equal to 0.00000, also, the magnitude of each charge are unexpectedly large. These cause big differences in link scheme calculation.

The result indicates that the settings and parameters, such as the number of grid points and layers in the approximation scheme are very different in both software packages. Hence, the only (currently) available QM software for the package is Gaussian. In order to include Turbomole and ORCA in this package, the implementation of RESP is essential. The solution of this situation is to write another small package of the RESP approximation, to restore the atomic charge at the same standard. Unfortunately, this task is beyond the scope of the present thesis.

Table 4.1.: RESP charge in Gaussian and Turbomole at BP86/STO-3G level, the testing model is GlyAla with the partitioning 1 (QM1-7) and link atom (hydrogen atom), (Unit: a.u.).

Atom	Gaussian	Turbomole
1 N	-0.166090	4.177669
2 H	0.303007	-1.433730
3 H	0.320549	-1.443164
4 H	0.292940	-1.102943
5 C	0.003268	7.102171
6 H	0.091428	-1.803344
7 H	0.111106	-2.250757
8 H	0.043792	-3.245902
sum of charges	1.00000	0.00000

4.1.3. gmx2qmmm output files

Energies and forces are the main result from hybrid QM/MM calculations. Thus, in addition to the *logfile* that prints out the operations in each time step of the QM/MM calculation, the output files of the **gmx2qmmm** are designed to contain the corresponding energy and forces.

Single point calculations

Following a single point calculation, output files include *oenergy.txt* and *oforce.txt*. The *oenergy.txt* contains the energy composition in details (QM energy, MM energy, link energy, and total energy). On the other hand, the *oforce.txt* displays the atomic forces with Cartesian coordinates and corresponding atom labels. The *oenergy.txt* and *oforce.txt* details are shown in Table 4.2.

<i>oenergy.txt</i>				
Step: 0				
QM energy:	-136.75602154			
MM energy:	0.09646171			
Link energy:	157.28624152			
Total energy:	-293.94580136			
<i>oforce.txt</i>				
1	-0.00055454	0.00038997	-0.00291820	
2	0.00010326	0.00353586	-0.00671161	
3	0.00052029	0.00514949	0.00276866	
4	-0.00033922	0.00454526	-0.00217659	
5	-0.00146200	-0.01437217	-0.00499166	
6	0.00631566	-0.00975389	0.00181975	
7	-0.00868157	-0.00318412	0.00343593	
8	-0.00032349	-0.01049729	-0.01040199	
9	-0.00355717	-0.01403005	-0.01426488	
10	-0.00243358	-0.00304776	-0.00466911	
11	0.00473331	0.00260438	-0.00058318	
12	-0.00342527	-0.00113160	0.00138708	
13	-0.00170051	-0.00094938	0.00818802	
14	0.00331946	-0.00083638	-0.00161156	
15	-0.00011167	0.00261732	0.00325083	
16	0.00744555	-0.00640143	0.00058721	
17	-0.00093010	0.00243895	-0.00609154	
18	-0.00533710	0.00443787	-0.00515468	
19	0.00612207	-0.00147283	0.00566980	
20	-0.00477500	0.00063397	-0.01219691	

Table 4.2.: Output files in single point calculation: *oenergy.txt* contains QM, MM, Link, Total energy; *oforce.txt* shows atomic forces for each atom with corresponding labels in Cartesian system; the first column stands for the atoms labels, the second, third and forth columns are the force in x,y,z component, respectively. (Unit: a.u.)

Geometry optimization

Output files include *oenergy.txt*, *opt_E.txt*, and *optforce.txt* in geometry optimization. The *oenergy.txt* prints the energy minimization, without listing the energies of steps that were rejected during the optimization process. To investigate the energy minimization in terms of all actually computed steps, *opt_E.txt* are designed to print out the energy composition

4. Results and Discussion

(QM, MM, Link, Total energy) for each actually computed single point. Also, the atomic forces from each single point calculation are printed out via *optforce.txt*. The output files that are relevant for a geometry optimization are shown in Table 4.3.

<i>oenergy.txt</i>				
1	-119.14964460			
2	-119.15920926			
3	-119.17356349			
4	-119.18730604			
5	-119.19375943			
<i>opt_E.txt</i>				
0	-96.25283588	0.12557041	78.65541827	-174.78268375
1	-96.25357445	0.11899580	78.65478075	-174.78935940
2	-96.25377802	0.11722351	78.65466131	-174.79121582
3	-96.25401668	0.11516191	78.65451936	-174.79337414
4	-96.25427100	0.11278424	78.65435667	-174.79584343
5	-96.25458400	0.11005279	78.65415511	-174.79868631
6	-96.25495166	0.10694982	78.65398570	-174.80198753
7	-96.25538009	0.10346102	78.65372092	-174.80564000
<i>optforce.txt</i>				
Step 1				
1	-0.00273027	0.00383569	-0.00074287	
2	-0.00177138	-0.01130215	-0.01470269	
3	-0.00968721	0.02051484	-0.00162368	
4	0.01898796	0.01270583	-0.00511922	
5	0.01548365	0.05190975	0.00076266	
6	0.05180788	-0.05844466	0.01900195	
7	-0.07558878	-0.02173128	0.01000368	
8	-0.00908435	-0.01954333	-0.02064631	
9	-0.00044030	-0.02203589	-0.01989641	
10	0.000345406	-0.01185213	-0.01511559	
11	0.000205806	0.00598820	0.00518760	
12	-0.00796471	0.00808913	-0.04435918	
13	0.00357575	-0.00942613	0.04810539	
14	-0.02591114	-0.03970939	0.00660087	
15	0.01053000	0.04436783	0.03631103	
16	0.03980939	-0.02971365	0.00064643	
17	-0.00721378	0.02222825	-0.04835557	
18	0.03788176	-0.09759898	0.22886736	
19	0.00639008	-0.00750388	0.01768671	
20	-0.05606094	0.12025216	-0.24538274	
Step 2				
1	-0.00245546	0.00413388	-0.00118293	
2	-0.00174398	-0.01101181	-0.01438379	
3	-0.00947523	0.02027223	-0.00161516	
4	0.01853846	0.01255686	-0.00515986	
5	0.01431156	0.04815525	0.00098930	
6	0.05017410	-0.05647478	0.01850902	
7	-0.07309425	-0.02083228	0.00993892	
8	-0.00852365	-0.01866930	-0.02035604	

Table 4.3.: Output files in geometry optimization: *oenergy.txt* contains the total energy without the rejected energies in energy minimization; *opt_E.txt* prints out all energies from all computed steps (QM, MM, Link, Total energy), the first column stands for the step order, the second to the fifth columns are QM, MM, Link, Total energy in each true step, respectively; *oforce.txt* shows atomic forces for each atom with corresponding labels in Cartesian system in each step. (Unit: a.u.)

4.2. Single point calculation

4.2.1. Energies result

Both GlyAla and GlySer and their rotated conformers are included in the comparison of the QM/MM energies. Their energies are calculated from Eq 3.1. We know that if the QM/MM calculation includes more QM atoms in QM calculation, the energy is lower due to the different definition of the zero point in the QM and MM potentials. Each partitioning contains different number of QM atoms. Therefore, from the aspect of energy, we are more interested in energy differences than absolute energies.

The errors in the subsection are all calculated by

$$E_{error} = \frac{E_{actual} - E_{fullQM}}{E_{fullQM}} \% \quad (4.1)$$

Due to steric repulsion effects, we can expect the energy after the rotation of the methyl group ($C_{(14)}, H_{(15)}, H_{(16)}$ and $H_{(17)}$) in GlyAla to be higher than in the original structure. More specifically, the hydrogen atoms ($H_{(13)}$ and $H_{(15)}$) are closer than before. As the result in Table 4.4, the hypothesis is correct; the energies differences are positive with full-QM calculation. Similarly in GlySer, the rotation of hydroxymethyl group ($C_{(14)}, H_{(15)}, H_{(16)}$, $O_{(17)}$ and $H_{(18)}$) causes a higher energy due to the steric effect. The steric effect is not only resulted from the hydrogen atoms ($H_{(13)}$ and $H_{(15)}$) as in GlyAla, but also from oxygen atoms ($O_{(17)}$ and $O_{(20)}$). Therefore, the energy differences in GlySer are higher.

We can observe that the results from the QM/MM program have the same sign as the results in full-QM calculation in both GlyAla and GlySer, which points out the model in the program have the ability to reproduce the qualitatively correct energy differences even with different QM and MM regions. Moreover, in spite of the result with small basis set, the errors in GlyAla are all between -50 and 50 %, which points out the approximated model performs well in the GlyAla case with different partitioning. However, although the program enables to judge the higher energy after the rotation, the values of the energies are overestimated; most errors in GlySer are larger than 100 % with the plus sign. This is possibly due to the oxygen in the hydroxymethyl group. Besides the steric effect from $H_{(13)}$ and $H_{(15)}$, the $O_{(17)}$ and $O_{(20)}$ have relatively high local negative charge. Thus, they have more influence to each other and the charge distribution in the system. In full-QM calculation, the atomic charges are localized by considering the polarizability of every atoms in the system. However, the QM/MM calculation does not contain the relocalization of the charge at the $O_{(20)}$ and $O_{(21)}$ atom affected by $O_{(17)}$, because the $O_{(20)}$ and $O_{(21)}$ are treated at the MM region. The $O_{(20)}$ and $O_{(21)}$ atomic charge are both -0.8055 in QM/MM calculation. On the other hand, the $O_{(20)}$ atomic charge becomes -0.656240 and -0.647909, and $O_{(21)}$ atomic charge becomes -0.810997 and -0.774558 by RESP method before and after the rotation, respectively. This is the disadvantage mentioned in electrostatic embedding (See Section 2.1.3); the

4. Results and Discussion

polarization in the MM region is not considered. Hence, the overestimation of energy differences occurs in GlySer cases.

Table 4.4.: GlyAla energy differences (after minus before rotation) with different partitioning at B3-LYP level (Unit: kJ/mol): the number in the parenthesis represents the errors which are calculated from Eq 4.1.

Basis	QM1-7	QM1-7,14-17	QM14-17	QM(full)
STO-3G	14.978(41.81%)	0.433(-95.90%)	14.656(38.76%)	10.562
3-21G	14.959(2.64%)	10.146(-30.39%)	16.539(13.48%)	14.574
6-31G*	14.956(11.16%)	11.506(-14.48%)	17.298(28.57%)	13.455
cc-pVTZ	14.923(16.40%)	15.243(18.89%)	18.043(40.73%)	12.821

Table 4.5.: GlySer energy differences (after minus before rotation) with different partitioning at B3-LYP level (Unit: kJ/mol): the number in the parenthesis represented the errors which are calculated from Eq 4.1.

Basis	QM1-7	QM1-7,14-18	QM14-18	QM(full)
STO-3G	93.855(262.41%)	51.933(100.53%)	99.861(285.61%)	25.897
3-21G	92.689(183.66%)	87.506(167.80%)	129.580(296.55%)	32.676
6-31G*	92.491(144.45%)	46.034(21.66%)	116.557(208.05%)	37.837
cc-pVTZ	90.548(155.69%)	35.297(-0.33%)	109.346(208.78%)	35.413

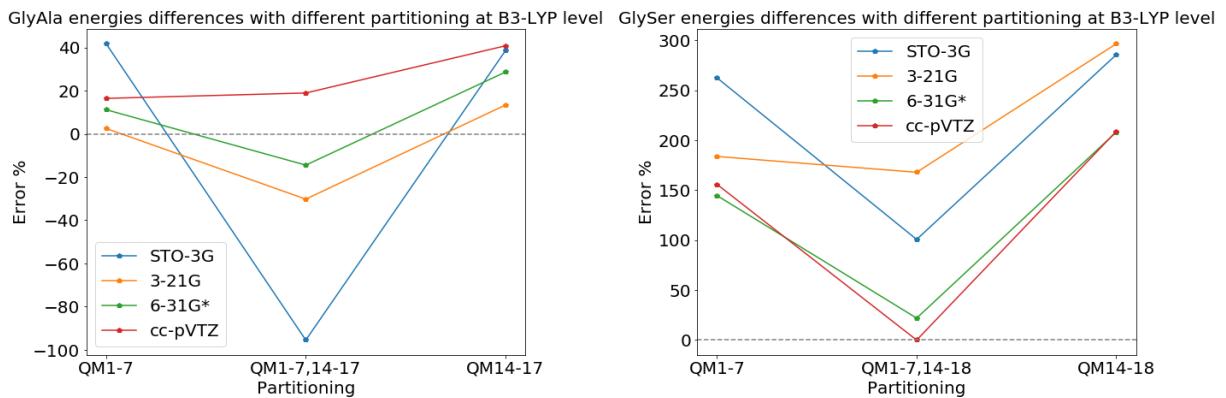


Figure 4.2.: Graph of energy difference with different partitioning at B3-LYP level, the grey dotted line stands for 0.00% error.

4.2. Single point calculation

Furthermore, the differences in energy upon rotation in the model systems can be analyzed further due to the data presented in Table 4.6 and 4.7 as well as the visualization in Figure 4.3 and 4.4, presenting the decomposition into the individual contributions.

In GlyAla, the result in partitioning 1 (QM1-7) well explains the energy difference is mainly from the MM energy, because the rotated methyl group locates in MM region. In partitioning 3 (QM14-17), which the rotated methyl group locates in QM region, the MM energy difference is still the main reason. But the QM energy difference becomes influential. However, the QM energy after the rotation is lower. The QM/MM model cannot well describe the polarization from the point charges field in MM region. Further, it appears that the approximated link energy corrections needs to be redesigned. The summation of the QM and MM energies (11.611 kJ/mol) at B3-LYP/6-31G* level is closer to the result in full-QM calculation (13.455 kJ/mol) than the one with the correction of link energy (17.298 kJ/mol). Thus, the link energy correction here is actually not needed. The energy resulting from partitioning 2, which contains both QM regions of partitioning 1 and 3, is not the additive result of them. One of the reasons is possibly that the point charges field after the charge shift scheme is different; the $N_{(10)}$ is a M_2 atom for both QM regions, therefore it contains two pairs of dipole-correcting point charges. Another reason might be the interaction of the two spatially separate QM regions. However, the QM energy difference is more influential than for partitioning 3. Again, the correction link energy is not effective to the result in this case.

On the other side, the results in GlySer are further away from the results in the full-QM system. First of all, we can observe that the overestimation of energy difference are mainly from MM energies. Further, the QM energy differences are all positive which increases the overestimation. The link energy scheme is not able to improve the results enough. The only difference between GlyAla and GlySer is the difference in rotated group, methyl and hydroxymethyl group. However, the results are very different. Therefore we can conclude, the approximated model in the program has large dependence on the conformation. The correction of the link scheme is not adapted to compensate for errors in the polarization. A potential solution to these problems is to improve the correction potential or to set up a more specific potential tailored for the current chemical situation.

4. Results and Discussion

Table 4.6.: GlyAla energy differences as their individual components with different partitioning at B3-LYP/6-31G* level; Total energy = QM energy + MM energy - Link energy (Unit: kJ/mol)

Level	QM1-7	QM1-7,14-17	QM14-17
QM energy	0.191	-5.965	-5.770
MM energy	14.770	17.389	17.381
Link energy	0.005	-0.0825	-5.688
Total energy	14.956	11.506	17.298

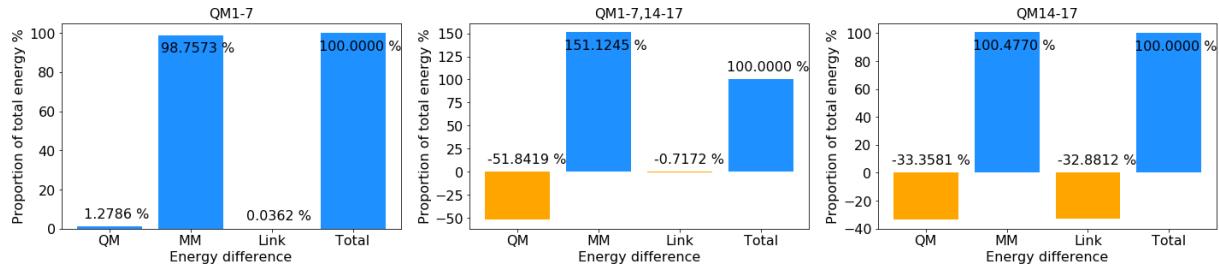


Figure 4.3.: GlyAla energy differences as their individual components (%) with different partitioning at B3-LYP/6-31G* level: The proportion is calculated as energy component/total energy (%); the blue bar and orange bar stand for positive and negative proportion, respectively.

Table 4.7.: GlySer energy differences as their individual components with different partitioning at B3-LYP/6-31G* level; Total energy = QM energy + MM energy - Link energy (Unit: kJ/mol)

Level	QM1-7	QM1-7,14-17	QM14-17
QM energy	19.259	26.803	24.209
MM energy	73.587	48.301	48.691
Link energy	0.355	29.071	-43.658
Total energy	92.491	46.034	116.557

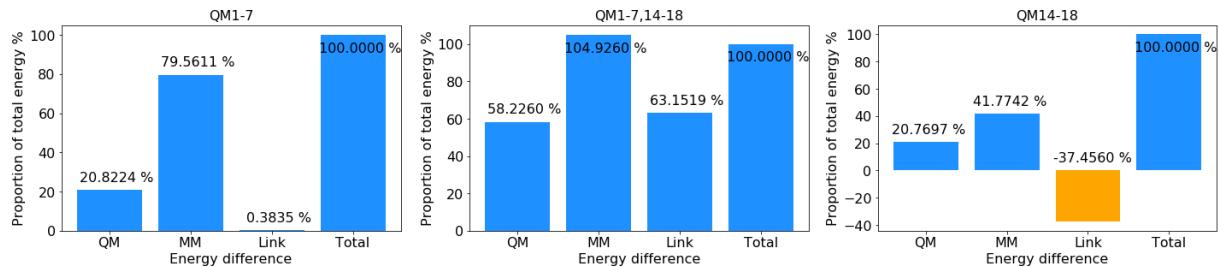


Figure 4.4.: GlySer energy differences as their individual components (%) with different partitioning at B3-LYP/6-31G* level: The proportion is calculated as energy component/total energy (%); the blue bar and orange bar stand for positive and negative proportion, respectively.

4.2.2. Forces result

The test model only contains the GlyAla (before rotation) in this section. This section includes the discussion and results of bond, angle and dihedral forces for the test system as well as the link bond forces.

Forces

The primary forces output in **gmx2qmmm** is the atomic forces in Cartesian coordinates. To investigate the forces between atoms, such as bond forces, angle forces and dihedral forces, the GF method is applied on the forces output; the forces matrix is multiplied by Wilson B matrix [49]. As a result, tables are produced containing $N - 1$ bond forces, $N - 2$ angle forces and $N - 3$ dihedral forces, where N is the number of atoms in the system. The force examples in Table 4.8 are the comparison of the forces between QM and MM regions and the illustration of errors is shown in Figures 4.5 and 4.6. The error is calculated as

$$F_{error} = \frac{F_{actual} - F_{fullQM}}{F_{fullQM}} \% \quad (4.2)$$

First of all, the forces results have the same sign as the results in the full-QM calculation. This points out the directions of the forces are all evaluated correctly. The bond force examples which are obtained from the QM calculation are maintained in an acceptable tolerance; errors are all between -100% and 100%. Also the calculation including more QM atoms has averagely accurate result. Especially for the bond forces in methyl group, the errors are between -10% and 10%.

The errors in angle and dihedral forces are mostly higher but still acceptable. Likely due to the accumulation of errors in atomic forces (more involved atoms for angle coordinates than for bond coordinates), some forces are overestimated in partitioning 2 (QM1-7,14-17). The overestimation takes place in the $H_{(3)}-N_{(1)}-H_{(2)}$ angle force and $H_{(4)}-N_{(1)}-H_{(3)}-H_{(2)}$ dihedral force. These two both contain atoms $N_{(1)}$, $H_{(2)}$ and $H_{(3)}$. This probably results from the additional pair of dipole correction charges on $N_{(10)}$ that affect the electrostatic force more. Therefore, because the whole system have to balance this bias, these two forces that contain nitrogen ($N_{(1)}$), which has a comparatively large electronegativity, may have comparatively large errors in partitioning 2.

4. Results and Discussion

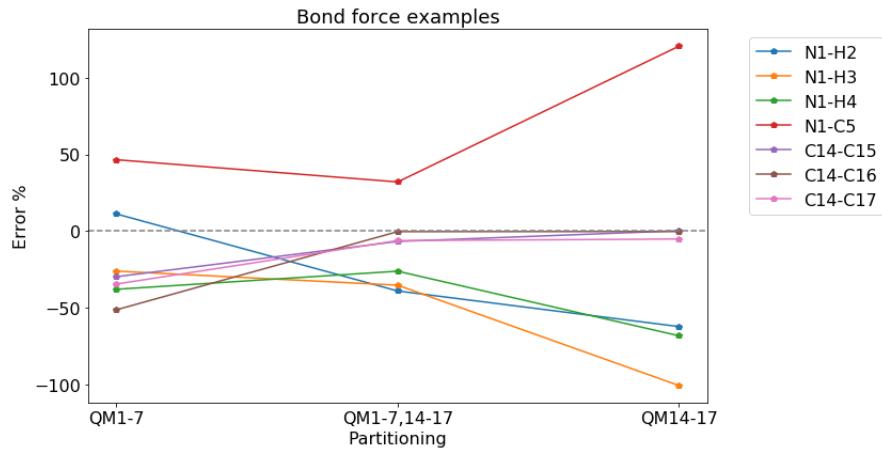


Figure 4.5.: Error of the bond forces in the example of GlyAla at B3-LYP/6-31G* level, the grey dotted line stands for 0.00% error.

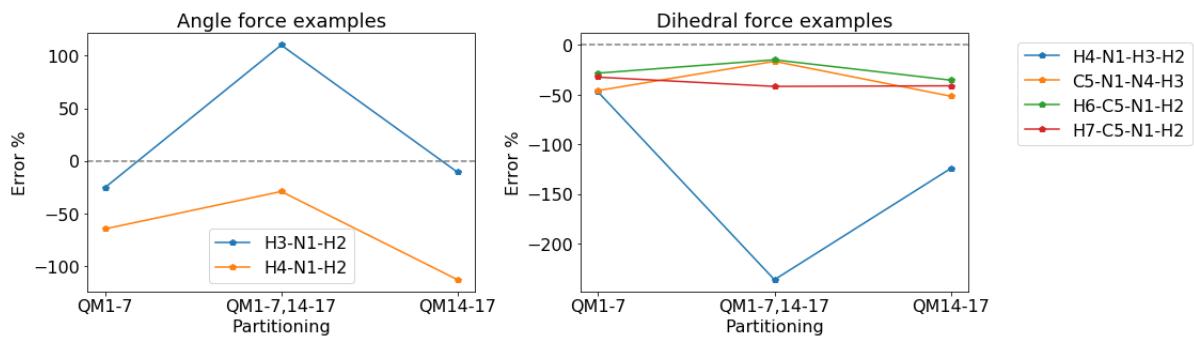


Figure 4.6.: Error of the angle (left) or dihedral (right) forces in the example of GlyAla at B3-LYP/6-31G* level, the grey dotted line stands for 0.00% error.

4.2. Single point calculation

Table 4.8: GlyAla force examples at the B3-LYP/6-31G* level, the MM calculation is at AMBER99 level (Unit: $kJ/(mol * a_0)$): the force examples include bond, angle and dihedral forces, the values in the parenthesis represent the errors which are calculated from Eq 4.2; in case the investigated coordinate is in a QM region, the force in printed as bold.

Bond	MM(full)	QM1-7	QM1-7,14-17	QM14-17	QM(full)
N ₍₁₎ -H ₍₂₎	18.24851289	53.254(11.22%)	29.2222(-38.97%)	18.125(-62.15%)	47.882
N ₍₁₎ -H ₍₃₎	-0.442	44.685(-26.04%)	39.146(-35.20%)	-0.334(-100.55%)	60.415
N ₍₁₎ -H ₍₄₎	31.811	62.096(-37.87%)	73.881(-26.08%)	31.845(-68.14%)	99.944
N ₍₁₎ -C ₍₅₎	-69.053	-45.715(46.52%)	-41.197(32.04%)	-68.787(120.47%)	-31.199
C ₍₁₄₎ -H ₍₁₅₎	232.961	233.055(-29.68%)	309.153(-6.71%)	332.245(0.25%)	331.406
C ₍₁₄₎ -H ₍₁₆₎	117.959	118.076(-51.22%)	241.304(-0.31%)	241.505(-0.23%)	242.062
C ₍₁₄₎ -H ₍₁₇₎	185.121	185.087(-34.43%)	264.889(-6.16%)	267.779(-5.13%)	282.270
Angle	MM(full)	QM1-7	QM1-7,14-17	QM14-17	QM(full)
H ₍₃₎ -N ₍₁₎ -H ₍₂₎	12.967	10.796(-25.37%)	30.399(110.14%)	12.916(-10.72%)	14.466
H ₍₄₎ -N ₍₁₎ -H ₍₃₎	3.397	-9.443(-64.61%)	-18.930(-29.05%)	3.448(-112.92%)	-26.683
Dihedral	MM(full)	QM1-7	QM1-7,14-17	QM14-17	QM(full)
H ₍₄₎ -N ₍₁₎ -H ₍₃₎ -H ₍₂₎	1.879	53.254(-47.39%)	29.2222(-236.56%)	18.125(-124.39%)	47.882
C ₍₅₎ -N ₍₁₎ -H ₍₄₎ -H ₍₃₎	35.615	44.685(-45.85%)	39.146(-16.40%)	-0.334(-51.54%)	60.415
H ₍₆₎ -C ₍₅₎ -N ₍₁₎ -H ₍₂₎	-73.943	62.096(-28.27%)	73.881(-14.87%)	31.845(-35.36%)	99.944
H ₍₇₎ -C ₍₅₎ -N ₍₁₎ -H ₍₂₎	49.190	-45.715(-32.34%)	-41.197(-41.68%)	-68.787(-41.06%)	-31.199

4. Results and Discussion

Link Forces

There are up to two link bonds in the testing system, $C_{(5)}-C_{(8)}$ and $C_{(12)}-C_{(14)}$, depending on the chosen partitioning. The force correction of each is

$$F_{C_{(5)}-C_{(8)}} = -0.01411644 \quad (4.3)$$

$$F_{C_{(12)}-C_{(14)}} = -0.01388719 \quad (4.4)$$

in atomic units. The remaining error is therefore calculated as

$$\begin{aligned} Rem.error &= \frac{A - B}{A} \% \\ A &= F_{fullQM} - (F_{actual} + F_{corr}) \\ B &= -F_{corr} \end{aligned} \quad (4.5)$$

which shows the remaining error after applying the force correction potential. It is acceptable if the remaining error is located in the range -100% to 100%. The results are shown in Table 4.9 and 4.10, together with the comparison of the errors with different basis set, which are illustrated in Figure 4.7.

The link forces have the same direction as the result of the full-QM calculation. Moreover, we can observe that the calculation with larger basis set or more QM atoms has better performance in $C_{(5)}-C_{(8)}$ bond force. On the other side, the errors of $C_{(12)}-C_{(14)}$ bond forces are very high. The errors are all over 100%. Again, this is probably because of the localized charge of oxygens ($O_{(19)}$ and $O_{(20)}$). Furthermore, the link force correction only includes the bond force correction. It is probably necessary to include angle and dihedral forces as well. Additionally, the link force correction should be sensitive to the conformation, especially for the atoms with larger polarizability. Otherwise, the results are likely not reliable with certain conformations.

Table 4.9.: GlyAla link bond forces between $C_{(5)}-C_{(8)}$ (Unit: $kJ/(mol * a_0)$): the values in the parenthesis represents the remaining error calculated from Eq 4.5.

Level	QM1-7	QM1-7,14-17	QM14-17	QM(full)
B3-LYP/STO-3G	-180.183(-603.97%)	-113.773(160.61%)	-56.513(131.30%)	-211.981
B3-LYP/3-21G	-188.061(69.00%)	-105.549(-0.02%)	-56.206(401.60%)	-105.557
B3-LYP/6-31G*	-184.069(73.51%)	-100.843(34.64%)	-55.370(-229.86%)	-81.197
B3-LYP/cc-pVTZ	-185.225(76.71%)	-81.920(33.63%)	-55.362(-26.54%)	-63.136

Table 4.10.: GlyAla link bond forces between C₍₁₂₎-C₍₁₄₎ (Unit: $kJ/(mol * a_0)$): the values in the parenthesis represents the remaining error calculated from Eq 4.5.

Level	QM1-7	QM1-7,14-17	QM14-17	QM(full)
B3-LYP/STO-3G	-139.651(126.94%)	-188.389(142.09%)	-213.683(159.45%)	-311.472
B3-LYP/3-21G	-139.906(157.86%)	-150.353(169.36%)	-178.774(251.02%)	-239.377
B3-LYP/6-31G*	-140.208(141.00%)	-155.686(149.63%)	-182.119(177.53%)	-265.608
B3-LYP/cc-pVTZ	-140.389(150.42%)	-151.675(159.75%)	-179.592(210.1%)	-249.161

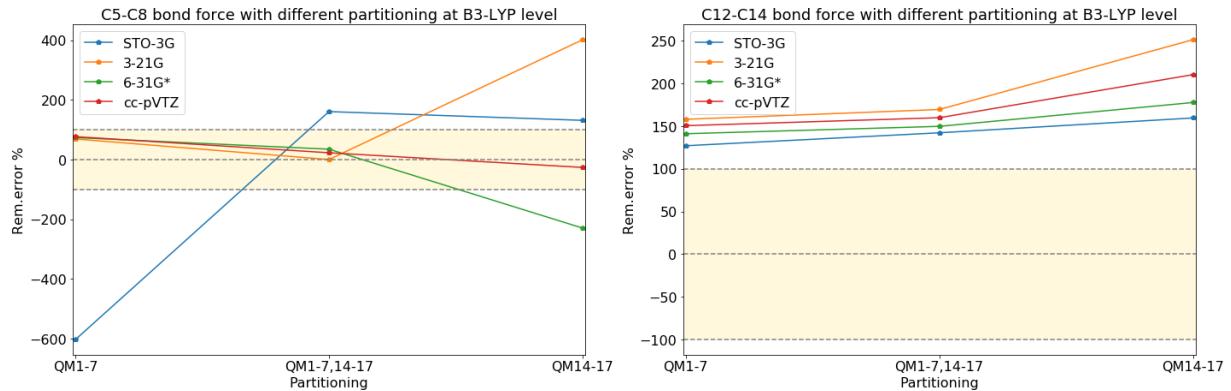


Figure 4.7.: The remaining error (%) after link bond force correction in GlyAla with different partitioning at B3-LYP level: the grey dotted line stands for 0.00% error, the bright yellow region stands for the acceptable remaining error between -100% to 100%.

4.3. Geometry optimization

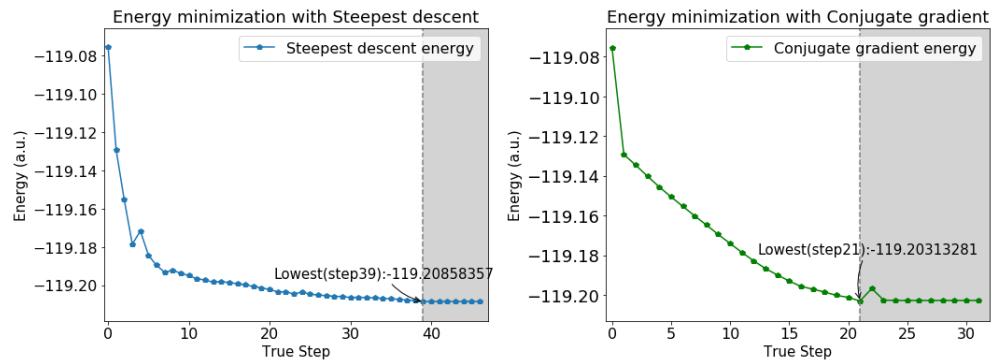
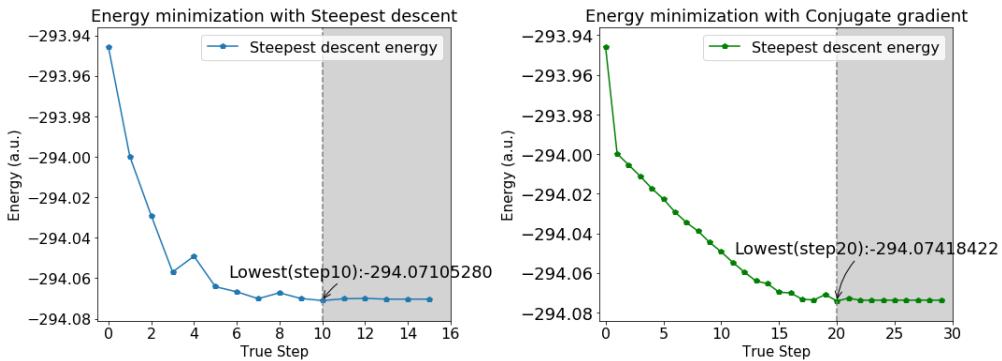
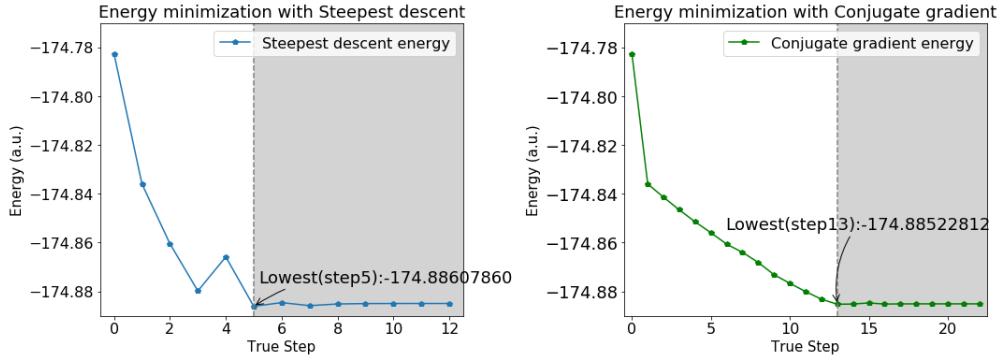
This section is about the geometry optimization result of the GlyAla before the rotation. The comparison of different algorithms and parameters, initial step size and update step size factor in the optimization are shown below.

4.3.1. Steepest descent vs Conjugate gradient

The Figures 4.8, 4.9 and 4.10 show the comparison of the steepest descent and conjugate gradient in partitioning 1, 2 and 3, respectively. The steepest descent searches for the direction of maximum forces, on the other hand, the conjugate gradient “remember” the information regarding the forces of previous steps, resulting in the searching direction containing a portion of previous direction. We find that the conjugate gradient is smoother during the energy minimization.

The Figure 4.11 shows the geometry obtained by geometry optimization in full-QM calculation and Figure 4.12 shows the geometry after the energy minimization in the

4. Results and Discussion



gmx2qmmm program. We can observe that the QM/MM optimization results are very different from the result obtained by full-QM calculation. In the full-QM calculation, the geometry and the bonds are changed; the hydrogen atom $H_{(3)}$ is connected to the oxygen atom $O_{(19)}$. However, the the QM/MM optimization results only have slight changes. The main reason is that the QM region does not contain the chemical reaction region, e.g. the oxygen $O_{(19)}$. The MM method is unable to describe describe process like bond breaking/formation.

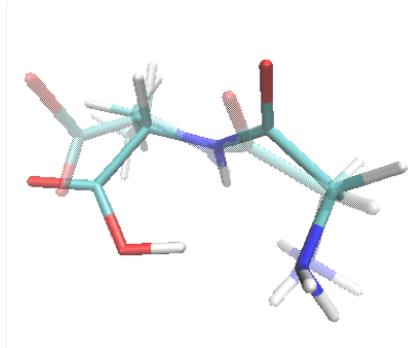


Figure 4.11.: GlyAla geometry after geometry optimization by full-QM calculation, the transparent part stands for the orginal structure of GlyAla at B3-LYP/6-31G* level.

In sum, the steepest descent and conjugate gradient is capable to run an optimization. However, it can be seen that the user should include the QM region where chemical processes take place in advance. Otherwise, we cannot obtain a result mimicking the result of the QM calculation due to the chemical processes not being included in the QM region.

4.3.2. Optimization with different initial step size

Besides algorithm in the optimizer, the initial step size is also a influential factor to the optimization with similar results. We can investigate the influence of different initial step sizes by the energy, maximum force and the number of computed steps in each calculation. The number of computed steps is the indicator of the run time of the calculation. The selected values for initial step size are 0.01, 0.05, 0.10 (default), 0.50 and 1.00 (a.u.). The results are illustrated in Figure 4.13 and 4.14, as well as the details shown in Table 4.11.

We can observe that the more iterations the calculation has, the lower energy and maximum force the result has in steepest descent. But in extreme case, such as the initial step size is 0.01 or 1.00 the running time is not positive correlated to the energy. The better result occurs with the initial step in the range between 0.05 to 0.1 in most cases in steepest descent.

In the conjugate gradient results, except the extreme cases, such as the initial step size is 0.01 or 1.00, the results are similar with different initial step size in partitioning 1

4. Results and Discussion

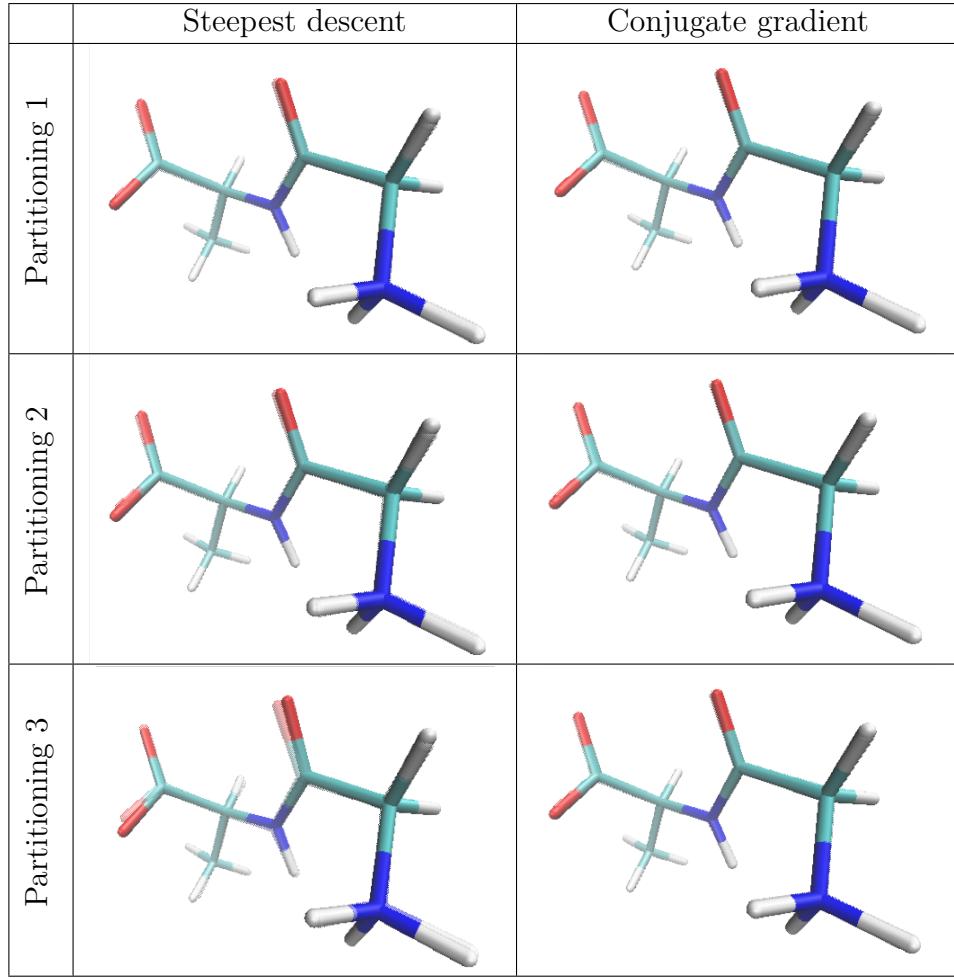


Figure 4.12.: GlyAla geometry with different algorithms and partitioning after geometry optimization, the transparent part stands for the initial geometry.

(QM1-7) and 2 (QM1-7,14-17). In the partitioning 3 (QM14-17), the result with the initial step size=0.05 has the lowest energy.

Table 4.11.: Optimization with different initial step size at B3-LYP/6-31G* level (Energy and force units: a.u.): successful steps and computed steps are abbreviated to S-steps and C-steps, respectively.

Propagator	Partitioning	Initial step size	S-steps	C-steps	Last energy	Max force
Steep	QM1-7	0.01	12	21	-174.8853	0.0218
Steep	QM1-7	0.05	6	15	-174.8873	0.0644
Steep	QM1-7	0.10	4	13	-174.8861	0.0991
Steep	QM1-7	0.50	6	16	-174.8869	0.0232
Steep	QM1-7	1.00	4	15	-174.8862	0.0343
Steep	QM1-7	2.00	6	17	-174.8867	0.0500

4.3. Geometry optimization

Steep	QM1-7,14-17	0.01	14	23	-294.0732	0.0140
Steep	QM1-7,14-17	0.05	6	15	-294.0751	0.0663
Steep	QM1-7,14-17	0.10	7	16	-294.0710	0.0330
Steep	QM1-7,14-17	0.50	7	17	-294.0706	0.0350
Steep	QM1-7,14-17	1.00	6	17	-294.0715	0.0165
Steep	QM1-7,14-17	2.00	11	23	-294.0744	0.0119
Steep	QM14-17	0.01	13	22	-119.1975	0.0333
Steep	QM14-17	0.05	7	16	-119.1972	0.0213
Steep	QM14-17	0.10	34	47	-119.2086	0.0051
Steep	QM14-17	0.50	34	48	-119.2087	0.0119
Steep	QM14-17	1.00	29	42	-119.2070	0.0068
Steep	QM14-17	2.00	5	16	-119.2040	0.1046
Conjgrad	QM1-7	0.01	39	51	-174.8862	0.0203
Conjgrad	QM1-7	0.05	18	28	-174.8853	0.0210
Conjgrad	QM1-7	0.10	13	23	-174.8852	0.0221
Conjgrad	QM1-7	0.50	13	24	-174.8852	0.0221
Conjgrad	QM1-7	1.00	23	36	-174.8862	0.0204
Conjgrad	QM1-7	2.00	15	27	-174.8859	0.0214
Conjgrad	QM1-7,14-17	0.01	31	42	-294.0744	0.0112
Conjgrad	QM1-7,14-17	0.05	23	34	-294.0748	0.0109
Conjgrad	QM1-7,14-17	0.10	19	30	-294.0742	0.0114
Conjgrad	QM1-7,14-17	0.50	19	31	-294.0742	0.0114
Conjgrad	QM1-7,14-17	1.00	1	11	-294.0246	0.2758
Conjgrad	QM1-7,14-17	2.00	11	23	-294.0747	0.0112
Conjgrad	QM14-17	0.01	35	46	-119.2049	0.0059
Conjgrad	QM14-17	0.05	60	75	-119.2120	0.0046
Conjgrad	QM14-17	0.10	21	32	-119.2031	0.0060
Conjgrad	QM14-17	0.50	21	33	-119.2031	0.0064
Conjgrad	QM14-17	1.00	21	33	-119.2043	0.0105
Conjgrad	QM14-17	2.00	1	12	-119.1312	0.2761

As a result, the initial step size does influence the optimization. However, this influence is hard to specifically state as it varies between the chosen parameter and model. What is found is that extreme cases of initial step size, such as greater than 1.00 or smaller than 0.01 are not suitable in the program.

4. Results and Discussion

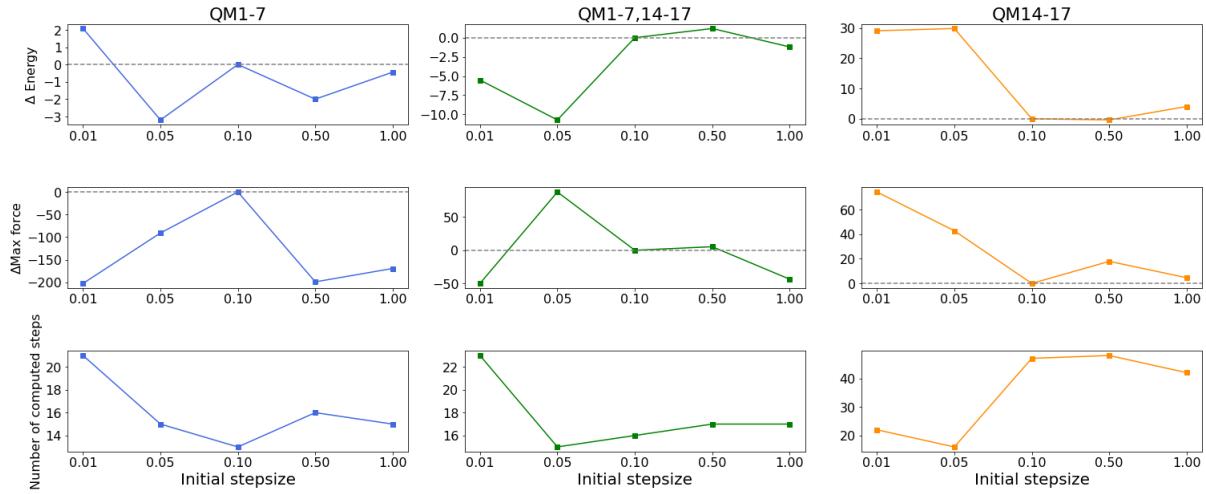


Figure 4.13.: Steepest descent with different initial step size at B3-LYP/6-31G* level (Energy unit: kJ/mol , Force unit: $kJ/(mol * a_0)$): the first row is the relative energy to the default value (initial step size = 0.1 a.u.), the second row is the relative max force to the default. The third row is the number of computed steps

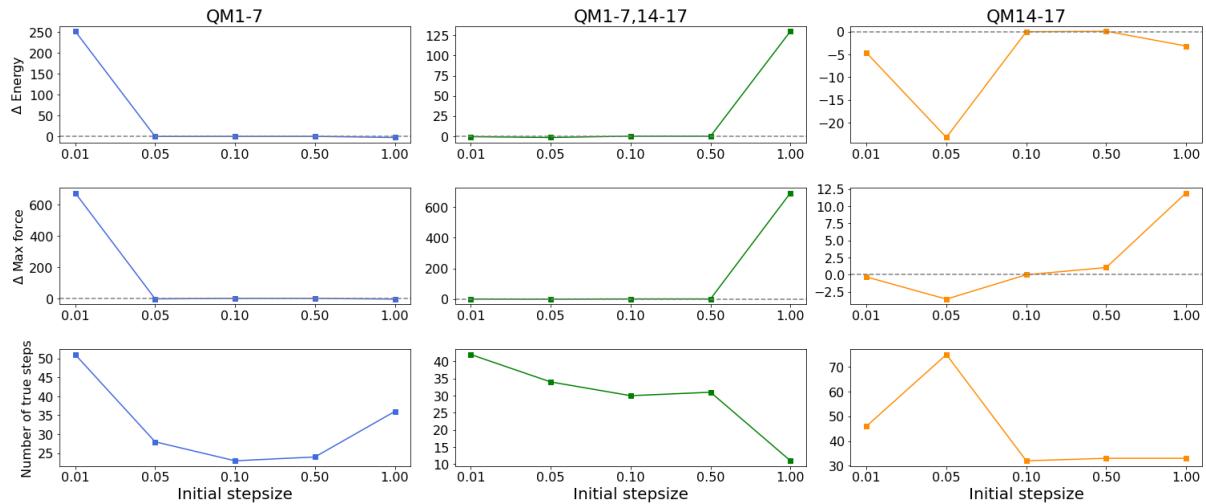


Figure 4.14.: Conjugate gradient with different initial step size at B3-LYP/6-31G* level (Energy unit: kJ/mol , Force unit: $kJ/(mol * a_0)$): the first row is the relative energy to the default value (initial step size = 0.1 a.u.), the second row is the relative max force to the default. The third row is the number of computed steps.

4.3.3. Optimization with different step size update factor

The step size update factor is mentioned in Section 3.5. The default value for the `gmx2qmmm` program is 0.2, which is the same setting in the GROMACS optimization. Because the GROMACS is an MM calculation software, therefore, for QM/MM calculation, it is interesting to apply different step size update factors and investigate the

4.3. Geometry optimization

relation to the result. The selected values for step size update factor are 0.05, 0.10, 0.20 (default), 0.30, 0.40 and 0.50. The results are illustrated in Figure 4.15 and 4.16, as well as the details shown in Table 4.12.

In steepest descent, mostly, the lowest energy is obtained from the step size update factor = 0.50. The larger the step size update factor is ,the longer is the run time. An acceptable range for the update factor appears to be between 0.10 and 0.30.

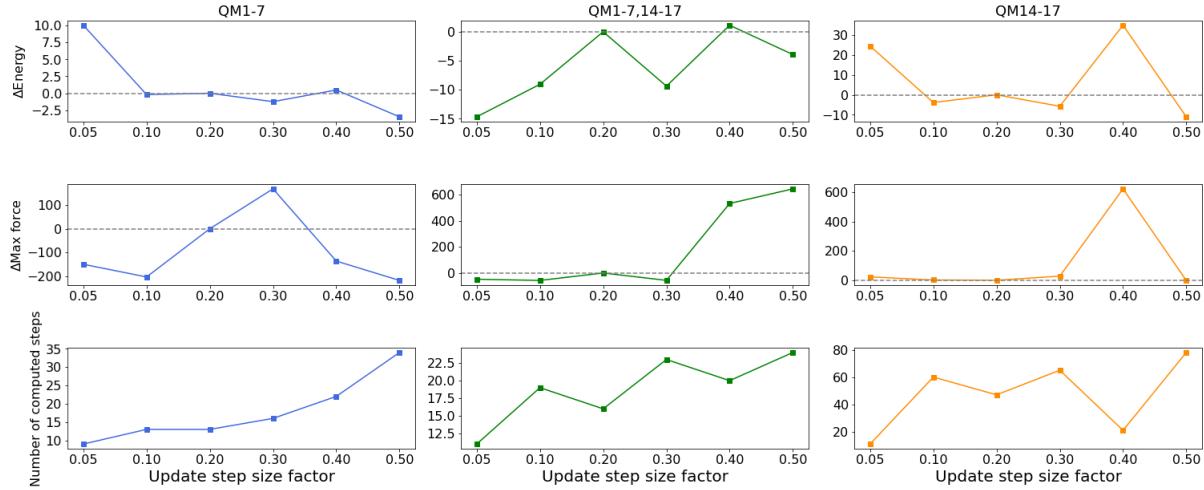


Figure 4.15.: Steepest descent with different step size update factors at B3-LYP/6-31G* level (Energy unit: kJ/mol , Force unit: $kJ/(mol * a_0)$): the first row is the relative energy to the default value (step size update factor = 0.2); the second row is the relative max force to the default; the third row is the number of computed steps.

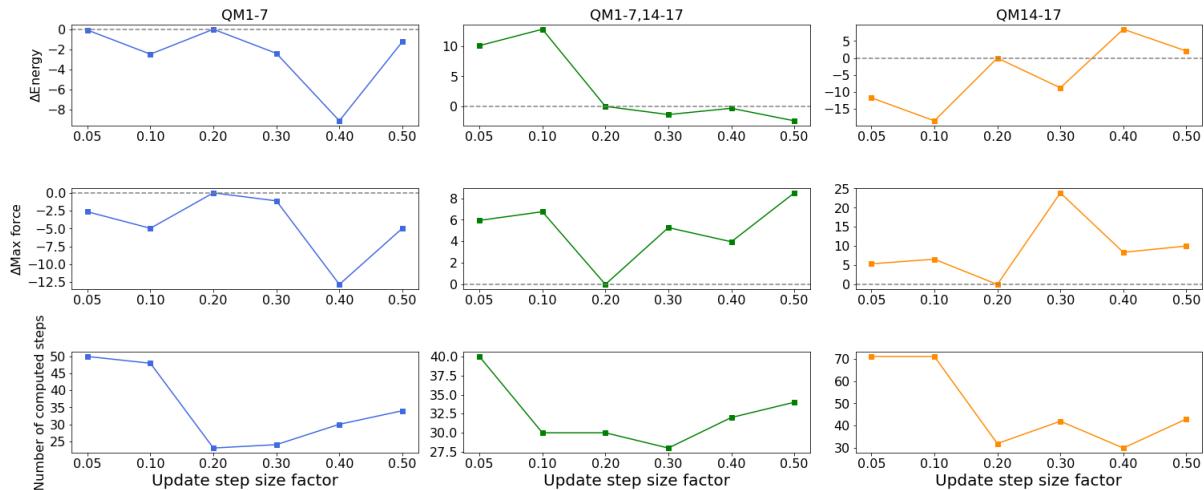


Figure 4.16.: Conjugate gradient with different step size update factors at B3-LYP/6-31G* level (Energy unit: kJ/mol , Force unit: $kJ/(mol * a_0)$): the first row is the relative energy to the default value (step size update factor = 0.2); the second row is the relative max force to the default; the third row is the number of computed steps.

4. Results and Discussion

Table 4.12.: Optimization with updated step size parameters at B3-LYP/6-31G* level (Energy and force units: a.u.): successful steps and computed steps are abbreviated to S–steps and C– steps, respectively.

Propagator	Partitioning	Initial step size	S-steps	C-steps	Last energy	Max force
Steep	QM1-7	0.05	4	9	-174.8822	0.0418
Steep	QM1-7	0.10	6	13	-174.8862	0.0214
Steep	QM1-7	0.20	4	13	-174.8861	0.0991
Steep	QM1-7	0.30	4	16	-174.8866	0.1633
Steep	QM1-7	0.40	6	22	-174.8859	0.0470
Steep	QM1-7	0.50	10	34	-174.8874	0.0158
Steep	QM1-7,14-17	0.05	6	11	-294.0766	0.0148
Steep	QM1-7,14-17	0.10	12	19	-294.0745	0.0113
Steep	QM1-7,14-17	0.20	7	16	-294.0710	0.0330
Steep	QM1-7,14-17	0.30	10	23	-294.0746	0.0116
Steep	QM1-7,14-17	0.40	4	20	-294.0706	0.2359
Steep	QM1-7,14-17	0.50	4	24	-294.0725	0.2785
Steep	QM14-17	0.05	6	11	-119.1993	0.0138
Steep	QM14-17	0.10	51	60	-119.2100	0.0058
Steep	QM14-17	0.20	34	47	-119.2086	0.0051
Steep	QM14-17	0.30	44	65	-119.2108	0.0158
Steep	QM14-17	0.40	5	21	-119.1954	0.2423
Steep	QM14-17	0.50	38	78	-119.2128	0.0046
Conjgrad	QM1-7	0.05	44	50	-174.8853	0.0211
Conjgrad	QM1-7	0.10	40	48	-174.8862	0.0203
Conjgrad	QM1-7	0.20	13	23	-174.8852	0.0221
Conjgrad	QM1-7	0.30	11	24	-174.8861	0.0217
Conjgrad	QM1-7	0.40	12	30	-174.8887	0.0173
Conjgrad	QM1-7	0.50	10	34	-174.8857	0.0203
Conjgrad	QM1-7,14-17	0.05	34	40	-294.0703	0.0136
Conjgrad	QM1-7,14-17	0.10	23	30	-294.0693	0.0139
Conjgrad	QM1-7,14-17	0.20	19	30	-294.0742	0.0114
Conjgrad	QM1-7,14-17	0.30	14	28	-294.0747	0.0134
Conjgrad	QM1-7,14-17	0.40	13	32	-294.0743	0.0129
Conjgrad	QM1-7,14-17	0.50	10	34	-294.0751	0.0146
Conjgrad	QM14-17	0.05	65	71	-119.2076	0.0080
Conjgrad	QM14-17	0.10	62	71	-119.2102	0.0085
Conjgrad	QM14-17	0.20	21	32	-119.2031	0.0060
Conjgrad	QM14-17	0.30	25	42	-119.2065	0.0151
Conjgrad	QM14-17	0.40	12	30	-119.1999	0.0092
Conjgrad	QM14-17	0.50	16	43	-119.2023	0.0098

Similar to the initial step size, the step size update factor affects the optimization,

and it is hard to quantify the correlation between the factor and results. It seems that the extreme cases with worse performance are not included in the test models. The default value of 0.2 appears to be a proper choice for the systems at hand.

4.4. Normal mode analysis

The normal mode analysis is implemented as the workflow in Section 3.1.3 that is based on the theory in Section 2.3. The result is shown in Table 4.13, which points out that the output is different from the result obtained in ORCA. Note that this test run includes only a given Hessian that is identical to the one provided for the ORCA calculation. Moreover, Table 4.13 points out the differences become larger with larger system. A possible reason for the larger errors are numerical problems caused by the Gram–Schmidt orthogonalization process, which generates the orthonormal matrix that removes the translation and rotation coordinates from the matrix. The process generates $3N - 6$ (the test models were all non-linear) random vectors with $3N$ elements. These random vectors are theoretically orthogonal to each other and the transformation matrix \mathbf{D} in Section 2.3 as well. Whereas, as the N increased, the resulting matrix \mathbf{D}' (after the Gram–Schmidt orthogonalization) is not a purely orthonormal matrix; the dot products of vectors in matrix \mathbf{D}' deviate from zero.

Table 4.13.: Vibrational frequencies (in cm^{-1}) for a small 18 atom test system (left) vs a 89 atom test system (right), computations were started from the input Cartesian Hessian matrix in all cases.

ORCA	gmx2qmmm	ORCA	gmx2qmmm
227.69	227.41369535	-36.81	-26.91375319
228.78	228.65289030	-19.88	-6.10165605
375.23	375.14175181	17.43	9.88188876
...
3004.29	2999.72171831	3217.15	2925.34151655
3016.57	3008.53731869	3273.95	2976.42778650

5. Summary and Outlook

The first part of the thesis is improving the implementation of **gmx2qmmm**. Aforementioned, the program should be re-organized in class structure and updated to the higher python version (at least python3). The main reason would be useful for data analysis are gradually available only in python3. Also, if we want to include the Turbomole and ORCA as QM software, the implementation of a RESP package is needed. This package should have the ability to restore the atomic charge with a proper fitting scheme.

The second part of the thesis is the justification of the approximated model in **gmx2qmmm**. We can conclude that most energies and forces results are acceptable with larger basis set. However, the results have large dependence on the conformation and the localized charge in the MM region. This is the a major drawback of the electrostatic embedding, which does not include the polarization in the MM region. The link energy and force correction should be able to compensate the error. However, corrections themselves rely on the conformation and the localized charge deeply, which are not accurate enough for certain cases. Besides using higher order fitting functions in the corrections, another potential improvement of this situation is to try more approximated models, with more amino acids, not only with butane as a crude basic model. For the different conformations, the program should be able to recognize the conformation then apply the similar or reasonable approximated model in the calculation.

The optimization includes steepest descent and conjugate gradient in the program. Several parameters such as initial step size and step size update factor affect the optimization. Unfortunately, their correlation to the quality of the results are unclear. Likewise, there exists more factors that influence theresults, for example, the threshold of the force in the iteration criteria. The main problem in the optimization is the discontinuity in the potential energy surface, which is the reason that we obtain a different minimum than for a full-QM calculation. The solution to this problem may be to use an interpolative region between two regions, creating a quasi continued potential surface [50].

The numerical problems in the normal mode analysis are the obstacle to analyze the molecular vibrations in QM/MM calculation. We have to find a numerically stable way to generate the orthonormal matrix. If the NMA is implemented, besides the vibrational analysis, the second order optimizer can be included in the program, then to overcome the limitations of the implemented optimizers.

Above all, the demand of more approximated model is for sure. And the correlation

5. Summary and Outlook

between the parameters in **gmx2qmmm** and the QM/MM calculation is unclear and entangled. It is unrealistic to justify the relation to the result for each parameter. Hence, machine learning may be a effective tool to solve the problems found for the QM/MM potential. The most difficult part for machine learning in the program is the quantity of the data; a better model obtained by the machine learning require enough data. Besides the quantity of data, the training model for the QM/MM calculation is intuitive; it is simple to determine the cost function. The main idea of the QM/MM method is generating the result that is close to full-QM calculation with reasonable time. Thus, the optimization in the training process is to find the possible set of parameters that reduce the difference between QM/MM and full-QM calculation.

Besides the determination of parameters in the program, the machine learning may have more application to QM/MM calculation [50]. Since the QM and MM calculation predict the energy in different ways, a discontinuity exists in the potential energy surface on the QM-MM boundary. A well-trained machine learning model can compensate the gap or define the interpolative region between the QM and MM regions. Then, we can have better result in the geometry optimization and other calculations..

All in all, the QM/MM method is a beautiful method to simulate large complex systems. The implementation is not easy but worthy and interesting. Although **gmxqmmm** has a lot of prospect to improve, it has ability to deal with some energy and force result already. Also, the potential to extend the functions and the further applications is large and worthwhile. Hence, with the further techniques and methods mentioned above, **gmx2qmmm** can be a promising tool in scientific computation in the future.

A. Appendix

A.1. Normal mode analysis in internal coordinates frame

The process to yield vibrational wavenumbers by internal coordinates normal mode analysis is according to Gaussian [11]. The first step is to translate the system to the center of mass to the origin, then determine the principal axes of inertia. The position center of mass is obtained by:

$$\mathbf{R}_{COM} = \frac{\sum_i^N m_i \mathbf{r}_i}{\sum_i^N m_i} \quad (\text{A.1})$$

where the sums are over N particles. The coordinates of the system can be shifted with

$$\mathbf{r}_{COMi} = \mathbf{r}_i - \mathbf{R}_{COM} \quad (\text{A.2})$$

to the center of mass. Next, the moments of inertia matrix \mathbf{I} is calculated as

$$\mathbf{I} = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix} = \begin{bmatrix} \sum_i^N m_i(y_i^2 + z_i^2) & -\sum_i^N m_i(x_i y_i) & \sum_i^N m_i(x_i z_i) \\ -\sum_i^N m_i(y_i x_i) & \sum_i^N m_i(x_i^2 + z_i^2) & -\sum_i^N m_i(y_i z_i) \\ -\sum_i^N m_i(z_i x_i) & -\sum_i^N m_i(z_i y_i) & \sum_i^N m_i(x_i^2 + y_i^2) \end{bmatrix} \quad (\text{A.3})$$

After the diagonalization of the inertia matrix \mathbf{I} ,

$$\mathbf{I}\mathbf{X} = \mathbf{X}I', \quad (\text{A.4})$$

it yields the principal moments (the eigenvalue I') and a 3×3 matrix (\mathbf{X}). Afterwards, the \mathbf{X} is used for generate the transformation \mathbf{D} to eliminate the translating and rotating coordinates. The first three vectors ($\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$) with $3N$ length are translational vectors. They defined as $\sqrt{m_i}$ times the corresponding coordinate axis

$$\begin{aligned} \mathbf{D}_1 &= (\sqrt{m_1}, 0, 0, \sqrt{m_2}, 0, 0, \dots, \sqrt{m_N}, 0, 0)^t \\ \mathbf{D}_2 &= (0, \sqrt{m_1}, 0, 0, \sqrt{m_2}, 0, \dots, 0, \sqrt{m_N}, 0)^t \\ \mathbf{D}_3 &= (0, 0, \sqrt{m_1}, 0, 0, \sqrt{m_2}, \dots, 0, 0, \sqrt{m_N})^t \end{aligned} \quad (\text{A.5})$$

A. Appendix

Another three vectors are rotational vectors, the element in these vectors are defined as

$$\begin{aligned} D_{4j,i} &= ((P_y)_i)X_{j,3} - (P_z)_i)X_{j,2})/\sqrt{m_i} \\ D_{5j,i} &= ((P_z)_i)X_{j,1} - (P_x)_i)X_{j,3})/\sqrt{m_i} \\ D_{6j,i} &= ((P_x)_i)X_{j,2} - (P_y)_i)X_{j,1})/\sqrt{m_i} \end{aligned} \quad (\text{A.6})$$

where $j = x, y, z$; i is the i^{th} atom over N atom and P is the dot product of the shifted coordinates and the corresponding row of matrix \mathbf{X} .

After the normalization of matrix \mathbf{D} , the Gram-Schmidt orthogonalization (See Appendix A.2) generates 3N-6 (or 3N-5 for a linear case) vectors, which are orthogonal to the translational and rotational vectors. The result is a transformation matrix \mathbf{D}' that transforms the mass weighted Cartesian coordinates \mathbf{q} to internal coordinates,

$$\mathbf{S} = \mathbf{D}'\mathbf{q} \quad (\text{A.7})$$

where the translation and rotation are removed.

At last, the mass-weighted Hessian matrix \mathbf{H}_{MWC} is transformed by \mathbf{D}' to internal coordinates Hessian matrix, \mathbf{H}_{INT} with diagolization of the $3N - 6$ (or 3N-5 in linear case) coordinates,

$$\mathbf{H}_{INT} = \mathbf{D}'^\dagger \mathbf{H}_{MWC} \mathbf{D}' \quad (\text{A.8})$$

The diagonalization of the internal coordinates Hessian matrix \mathbf{H}_{INT} then yields the $3N - 6$ (or 3N-5 in linear case) eigenvalues $\lambda = 4\pi^2\nu^2$, where ν is the vibrational wavenumber in normal mode analysis.

A.2. Gram–Schmidt method

Gram–Schmidt method [51] is named after Jørgen Pedersen Gram and Erhard Schmidt. It is a method to generate orthogonal set $S' = \{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ from a finite, linearly independent set $S = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ in a n -dimensional inner product space, where $k \leq n$.

In first step, let

$$\mathbf{y}_1 = \mathbf{x}_1, \mathbf{x}_1 \neq 0 \quad (\text{A.9})$$

Secondly, we can obtain \mathbf{y}_2 from the \mathbf{x}_2 after the subtraction of the \mathbf{x}_2 projection on \mathbf{y}_1 , which is

$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{\mathbf{y}_1^T \mathbf{x}_2}{\mathbf{y}_1^T \mathbf{y}_1} \mathbf{y}_1 \quad (\text{A.10})$$

and its simplified visualization is shown in Figure A.1.

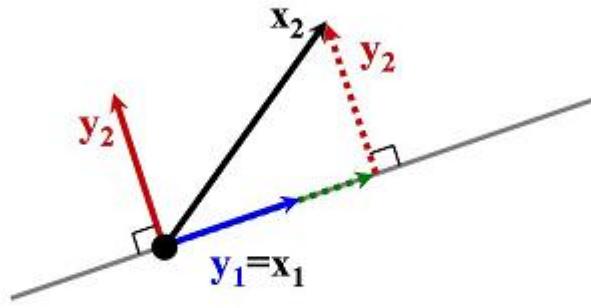


Figure A.1.: The first two steps of the Gram–Schmidt method, the figure is taken from reference [52]

Similarly, \mathbf{y}_3 is \mathbf{x}_3 after the subtraction of \mathbf{x}_3 projection on \mathbf{y}_1 and \mathbf{y}_2 , which is

$$\mathbf{y}_3 = \mathbf{x}_3 - \frac{\mathbf{y}_1^T \mathbf{x}_3}{\mathbf{y}_1^T \mathbf{y}_1} \mathbf{y}_1 - \frac{\mathbf{y}_2^T \mathbf{x}_3}{\mathbf{y}_2^T \mathbf{y}_2} \mathbf{y}_2 \quad (\text{A.11})$$

and its simplified visualization is shown in Figure A.2.

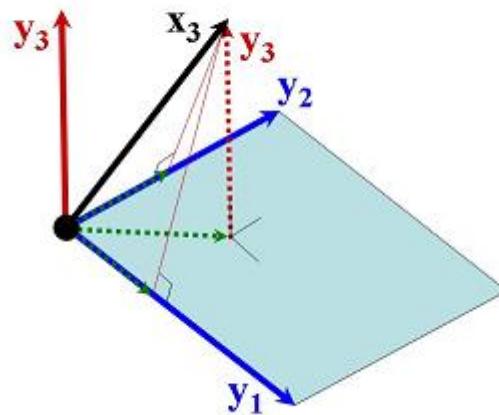


Figure A.2.: The third step of the Gram–Schmidt method, the figure is taken from reference [52]

A. Appendix

Therefore, \mathbf{y}_k can be expressed as

$$\mathbf{y}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\mathbf{y}_i^T \mathbf{x}_k}{\mathbf{y}_i^T \mathbf{y}_i} \mathbf{y}_i \quad (\text{A.12})$$

Bibliography

- [1] Ragnar Bjornsson and Michael Bühl. Modeling molecular crystals by qm/mm: Self-consistent electrostatic embedding for geometry optimizations and molecular property calculations in the solid. *Journal of Chemical Theory and Computation*, 8(2):498–508, 2012. PMID: 26596600.
- [2] Walter Thiel Hans Martin Senn. QM/MM methods for biomolecular systems. *Angewandte Chemie. International Edition.*, 48(7):1198–1229, 2009.
- [3] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227 – 249, 1976.
- [4] Alan R. Fersht. Profile of martin karplus, michael levitt, and arieh warshel, 2013 nobel laureates in chemistry. *Proceedings of the National Academy of Sciences*, 110(49):19656–19657, 2013.
- [5] National Research Council. *Mathematical Challenges from Theoretical/Computational Chemistry*. The National Academies Press, Washington, DC, 1995.
- [6] Gerrit Groenhof. *Introduction to QM/MM Simulations*, pages 43–66. Humana Press, Totowa, NJ, 2013.
- [7] Hai Lin and Donald G. Truhlar. Qm/mm: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts*, 117(2):185, Jul 2006.
- [8] Shruti Koulgi, Archana Achalere, Neeru Sharma, Uddhavesh Sonavane, and Rajendra Joshi. Qm-mm simulations on p53-dna complex: a study of hot spot and rescue mutants. *Journal of Molecular Modeling*, 19(12):5545–5559, Dec 2013.
- [9] Manuel Hitzenberger, Manussada Ratanasak, Vudhichai Parasuk, and Thomas S. Hofer. Optimizing link atom parameters for dna qm/mm simulations. *Theoretical Chemistry Accounts*, 135(3):47, Feb 2016.
- [10] Michael Levitt and Arieh Warshel. Extreme conformational flexibility of the furanose ring in dna and rna. *Journal of the American Chemical Society*, 100(9):2607–2613, 1978.

Bibliography

- [11] Gaussian 16 team. *Gaussian 16 user reference*.
- [12] Turbomole team. *Turbomole 7.3 manual*.
- [13] ORCA team. *ORCA 4.0.1 manual*.
- [14] GROMACS team. *GROMACS 2019 manual*.
- [15] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. Charmm-gui: A web-based graphical user interface for charmm. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.
- [16] David A. Case, Thomas E. Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz Jr., Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J. Woods. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688, 2005.
- [17] Jan Philipp Götze. *gmx2qmmm user reference*.
- [18] Yingkai Zhang, Tai-Sung Lee, and Weitao Yang. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *The Journal of Chemical Physics*, 110(1):46–54, 1999.
- [19] Visvaldas Kairys and Jan H. Jensen. Qm/mm boundaries across covalent bonds: A frozen localized molecular orbital-based approach for the effective fragment potential method. *The Journal of Physical Chemistry A*, 104(28):6656–6665, 2000.
- [20] Nathalie Reuter, Annick Dejaegere, Bernard Maigret, and Martin Karplus. Frontier bonds in qm/mm methods: A comparison of different approaches. *The Journal of Physical Chemistry A*, 104(8):1720–1735, 2000.
- [21] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A generalized hybrid orbital (gho) method for the treatment of boundary atoms in combined qm/mm calculations. *The Journal of Physical Chemistry A*, 102(24):4714–4721, 1998.
- [22] Guido van Rossum and Python team. *Python 2.7.16 documentation*.
- [23] Eric Jones Travis Oliphant, Pearu Peterson and Scipy team. *Scipy 1.3.1 documentation*.
- [24] Travis Oliphant and NumPy team. *NumPy v1.17 manual*.
- [25] John D. Hunter and Matplotlib team. *Matplotlib 3.1.1 documentation*.
- [26] Wes Mckinney and Pandas team. *Pandas v0.25.1 documentation*.

- [27] J. Espinosa-Garcia and J. C. Corchado. Reliability of the single-point calculation technique at characteristic points of the potential energy surface. *The Journal of Physical Chemistry*, 99(21):8613–8616, 1995.
- [28] H. Bernhard Schlegel. Geometry optimization. *Wiley Interdisciplinary Reviews*, 1(5):790–809, May 2011.
- [29] Ivet Bahar, Timothy R. Lezon, Ahmet Bakan, and Indira H. Shrivastava. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chemical Reviews*, 110(3):1463–1497, 2010. PMID: 19785456.
- [30] Marc W. van der Kamp and Adrian J. Mulholland. Combined quantum mechanics/-molecular mechanics (qm/mm) methods in computational enzymology. *Biochemistry*, 52(16):2708–2728, 2013. PMID: 23557014.
- [31] Lili Cao and Ulf Ryde. On the difference between additive and subtractive qm/mm calculations. *Frontiers in Chemistry*, 6:89, 2018.
- [32] Emmanuel M. Papamichael, Haralambos Stamatis, Panagiota-Yiolanda Stergiou, Athanasios Foukis, and Olga A. Gkini. Chapter 3 - enzyme kinetics and modeling of enzymatic systems. In Ram Sarup Singh, Reeta Rani Singhania, Ashok Pandey, and Christian Larroche, editors, *Advances in Enzyme Technology*, Biomass, Biofuels, Biochemicals, pages 71 – 104. Elsevier, 2019.
- [33] J.-M. Combes. The born-oppenheimer approximation. In Walter Thirring and Paul Urban, editors, *The Schrödinger Equation*, pages 139–159, Vienna, 1977. Springer Vienna.
- [34] Errol G. Lewars. *Ab initio Calculations*, pages 175–390. Springer Netherlands, Dordrecht, 2011.
- [35] Errol G. Lewars. *Density Functional Calculations*, pages 445–519. Springer Netherlands, Dordrecht, 2011.
- [36] Charlotte Froese Fischer. General hartree-fock program. *Computer Physics Communications*, 43(3):355 – 365, 1987.
- [37] Valerio Magnasco. Chapter 16 - post-hartree–fock methods. In Valerio Magnasco, editor, *Elementary Molecular Quantum Mechanics (Second Edition)*, pages 681 – 722. Elsevier, second edition edition, 2013.
- [38] L. H. Thomas. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(5):542–548, 1927.

Bibliography

- [39] E. Fermi. Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente. *Zeitschrift für Physik*, 48(1):73–79, Jan 1928.
- [40] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.
- [41] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.
- [42] Maylis Orio, Dimitrios A. Pantazis, and Frank Neese. Density functional theory. *Photosynthesis Research*, 102(2):443–453, Dec 2009.
- [43] P.E. Gill, W. Murray, and M.H. Wright. *Practical optimization*.
- [44] Jan A. Snyman and Daniel N. Wilke. *Line search descent methods for unconstrained minimization*, pages 41–69. Springer International Publishing, Cham, 2018.
- [45] E.B. Wilson. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. McGraw-Hill Interamericana, 1955.
- [46] Emil Zak. Normal modes. the true story, 2016.
- [47] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.
- [48] Brent H. Besler, Kenneth M. Merz Jr., and Peter A. Kollman. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry*, 11(4):431–439, 1990.
- [49] Lee-Ping Wang and Chenchen Song. Geometry optimization made simple with translation and rotation coordinates. *The Journal of Chemical Physics*, 144(21):214108, 2016.
- [50] Yin-Jia Zhang, Alireza Khorshidi, Georg Kastlunger, and Andrew A. Peterson. The potential for machine learning in hybrid qm/mm calculations. *The Journal of Chemical Physics*, 148(24):241740, 2018.
- [51] Åke Björck. Solving linear least squares problems by gram-schmidt orthogonalization. *BIT Numerical Mathematics*, 7(1):1–21, Mar 1967.
- [52] 周志成 . Gram–schmidt 正交化與qr 分解, April 2010.