

Computational Studies of Many-body effects in Molecular Crystals

Dissertation

for the award of the degree

„Doctor rerum naturalium“

of the Georg-August-Universität Göttingen

within the doctoral program Chemistry

of the Georg-August University School of Science (GAUSS)

submitted by

Thorsten Lennart Teuteberg

from Aurich

Göttingen, 07.12.2018

Thesis Committee

Supervisor:

Prof. Dr. Ricardo A. Mata
Computational Chemistry and Biochemistry
Institut für Physikalische Chemie

Second Supervisor:

Prof. Dr. Dietmar Stalke
Anorganische Molekül- und Strukturchemie
Institut für Anorganische Chemie

Examination Board

Reviewer:

Prof. Dr. Ricardo A. Mata
Computational Chemistry and Biochemistry
Institut für Physikalische Chemie

Second Reviewer:

Prof. Dr. Dietmar Stalke
Anorganische Molekül- und Strukturchemie
Institut für Anorganische Chemie

Further Members of the Examination Board:

Prof. Dr. Jörg Behler
Theoretische Chemie
Institut für Physikalische Chemie

PD Dr. Thomas Zeuch
Physikalische Chemie II
Institut für Physikalische Chemie

Prof. Dr. Götz Eckold
Physikalische Chemie fester Körper
Institut für Physikalische Chemie

Dr. Tim Schäfer
Physikalische Chemie I
Institut für Physikalische Chemie

Date of the oral examination: 25.01.2019

Contents

Acknowledgement	iii
Oath	iv
Abstract	v
1 Introduction	1
2 Theory	5
2.1 Wave-function Theory	5
2.1.1 The Hartree-Fock Method	6
2.1.2 Møller-Plesset Perturbation Theory	7
2.1.3 Coupled-Cluster Theory	8
2.1.4 Local Electron Correlation Methods	9
2.2 Density Functional Theory	10
2.2.1 Ranks of Density Functional Theory	10
2.2.2 Dispersion Corrections	11
2.3 Symmetry-Adapted Perturbation Theory	11
2.4 General Hybrid QM/MM	12
2.5 Electronic Structure Calculations of Molecular Crystals	13
2.5.1 Periodic Wave-function Approaches	13
2.5.2 Additive Approaches	14
3 Methods	19
3.1 Computational Details	19
3.2 Statistical Methods	20
3.3 The Benchmarking Reference	20
4 The ac-QM/MM Model	23
4.1 Bjornsson-Bühl QM/MM	23
4.2 Overview of the ac-QM/MM Scheme	25
4.3 A Consistent Energy Expression	27
4.4 Generalised Geometry Gradients and Crystal Optimisation	28
4.4.1 ac-QM/MM Internal Gradients	28
4.4.2 Cell Vector Gradients	35
4.4.3 Stress and Strain	39
4.5 Choice of the Embedding Region	51
4.6 Charge Update Procedure	53
4.6.1 Implementation of the Update Scheme	53
4.6.2 Convergence Threshold	55
4.6.3 Charge Damping	58
4.6.4 Charge Damping with Different Thresholds	62
4.6.5 Background Charge Overfitting	64
4.7 Two-body ac-QM/MM	67
4.7.1 Selection of the Two-body Region	67
4.7.2 General Formalism ac(2)-QM/MM	68
4.7.3 Variation ac(2-F)-QM/MM	68
4.7.4 Variation ac(2-I)-QM/MM	69
4.7.5 Performance of the Two-body Region Selection	69
4.7.6 Convergence Behaviour with the Two-body Region Size	71
4.7.7 Applicability of a Dimer Approach	73
4.7.8 D3, or not D3, that is the Question	77
4.7.9 Crystal Symmetry Restrictions	78
4.7.10 ac(2)-QM/MM Application	79
4.7.11 ac(2-F)-QM/MM Application	79
4.7.12 ac(2-I)-QM/MM Application	80

4.8	Suggested Models for ac-QM/MM	82
4.8.1	ac(1)-QM/MM	82
4.8.2	ac(2)-QM/MM	89
5	ac-QM/MM Benchmarks and Applications	93
5.1	Performance of the ac-QM/MM Models	93
5.1.1	ac(1)-QM/MM	93
5.1.2	ac(2)-QM/MM and ac(2-I)-QM/MM	94
5.2	Applications of the ac-QM/MM Model	99
5.2.1	An Unusual Piano-Stool Complex	99
5.2.2	Investigation of a Diselenide X-ray Structure	100
6	Summary	107
7	Outlook	111
	Bibliography	113

Acknowledgement

Mein besonderer Dank gilt ...

... meinen Eltern für ihre unschätzbare praktische und moralische und emotionale Unterstützung, insbesondere während meiner melancholischen Phasen ...

... Prof. Ricardo Mata, der mich wundersamerweise immer wieder zum Weitermachen motiviert hat, wenn scheinbar gar nichts funktionierte (also gefühlt die erste Hälfte der Promotion) und die „Herausforderungen“* immer zahlreicher wurden. Per aspera ad astra ...

... Prof. Dietmar Stalke für die zahlreichen Gelegenheiten, über den Tellerrand der Theorie hinaus zu blicken. Auch wenn diese „kleinen“ Projekte bisweilen rasch um viele „Das könnte man dann ja auch eben noch rechnen, oder?“ gewachsen sind ...

... Arne, Roman, Ingo und Christian für die unkomplizierte Zusammenarbeit beim Blick über den Tellerrand hinaus ...

... Julian, Sebastian, Matheus, Kris, Christian und Michaela für das freitagmittägliche Kontrastprogramm zu den „Herausforderungen“* im Büro. Insbesondere für die Erkenntnis, dass die Arbeitsbedingungen im 4. Stock IPC doch signifikant über dem Durchschnitt der Chemie liegen ...

... Mirko und Axel für das kostenlose Hintergrundunterhaltungsprogramm im Büro. Das hat zwar bisweilen die Produktivität messbar beeinträchtigt, aber oft auch meine Arbeitsmoral signifikant gestärkt ...

... Beate für die Erkenntnis, dass man doch nicht (völlig) verrückt sein muss, um regelmäßig mit dem Rad vom Nordcampus auf den Nikolausberg zu fahren ...

... Wiebke, Ruth, Esther, Niels, Maximilian und Paul für die Erfahrung, dass Lehre tatsächlich richtig viel Spaß machen kann. Macht's gut, und danke für den Kuchen!

*„Es gibt keine Probleme, nur Herausforderungen.“, R. Mata 2011 zu einem an Aceton verzweifelnden Bachelor-Studenten

Oath

I hereby declare that I have prepared this thesis all by myself, not used any sources or tools except for those explicitly stated, and marked all quotes, be it in literal or analogous form, accordingly. I declare that I have not submitted this thesis, or parts thereof, to this or any other university in the context of a failed examination.

Göttingen, the 07.12.2018

(Thorsten Lennart Teuteberg)

Abstract

A method for multiscale modeling of molecular crystals is presented - the additive crystal quantum mechanics/molecular mechanics model (ac-QM/MM). At the one-body level, a single molecule is chosen as the QM region. The surrounding MM region consists of a finite cluster of explicit MM atoms, represented by point charges and Lennard-Jones potentials, with additional background charges to mimic periodic electrostatics. QM-derived charges are obtained self-consistently to enable polarisation of the environment. An expansion with two-body corrections to the interaction of the central molecule with its nearest neighbours is also included. Fully analytical gradient expressions for both the molecular geometry and the cell parameters have been developed, hence, allowing for an unrestricted optimisation. Both the lattice energy and, in principle, all molecular properties can be calculated within this scheme.

Benchmarking the approach with the X23 reference set yields reasonable geometries and cohesive energies for many systems, but hydrogen bonded molecules pose a significant challenge. The inclusion of two-body corrections greatly improves the results. Differences to experimental findings are significant, but similar deviations are obtained across the whole X23 set. A comparison to plane-wave DFT reveals a quite systematic overestimation of cohesive energies by 6-16 kJ · mol⁻¹ proving a consistent behaviour of the ac-QM/MM model. The primary objective is to offer an inexpensive and flexible method to account for the crystal environment, but it can be converged to chemical accuracy if correlated wave function methods are applied. Due to the requirement of numerical gradients within the QM part these are only applicable to the smallest systems, but allow for impressive comparisons to experiment. Similar calculations based on DFT optimised geometries unfortunately do not provide overall improved results among the X23 set.

The ac-QM/MM model was applied to an interesting crystal structure, which shows an uncommon bending of electron-poor hydrogen atoms towards a cation. From electrostatics, a repulsion between hydrogen atoms (positive partial charge) and the cations would be expected. Gas phase calculations show this expected inclination of hydrogen atoms away from the cation. However, the ac-QM/MM scheme qualitatively reproduces the experimental bending.

The investigation of an experimentally challenging X-ray structure of a diselenium compound, providing distinct charge accumulations aside the selenium atoms, could also be aided. An ac-QM/MM study helped to determine a triplet state to be the most probable source of the charge concentrations. The respective structure was found to be mostly similar to the ground state geometry, but a broken Se-C bond leads to a distinctly different orientation of the Se atoms. These are likely responsible for the charge accumulations observed experimentally.

1 Introduction

The computation of molecular systems in real laboratory/life/reactor conditions is a grand challenge, one that depends on many different steps. For example, simulating a molecule also requires the description of its environment. However, most quantum chemical methods are built for the treatment of distinct systems.

In the lab molecules are found in a medium, be it solvent molecules in a solution, atoms or ions at a surface, or other molecules of the same kind in a molecular crystal. The fundamental concept of a molecule reflects the knowledge that covalent forces are stronger than the interactions with the environment and, thus, most important in defining its properties. Consequently, neglecting other contributions and restricting a calculation to a single molecule is frequently found to be a reasonable first approximation. The treatment of molecules in chemistry has made great advances and both increasing computational power and sophisticated methods allow for the calculation of increasingly large molecules. Single molecules with hundreds of atoms can already be optimised with density functional theory techniques, and energies for comparably large systems become more and more amenable with wave-function theory methods under appropriate approximations.¹

Nevertheless, the interactions of a molecule with its environment exist and contribute to the internal molecular conformation and its arrangement to other molecules. The interactions among the molecules in a molecular crystal not only allow for a single crystal structure, but for several different packing motifs. Often several of these polymorphs are very similar in energy and, hence, potentially accessible from crystallisation. Yet these differences may have significant impact on the properties of the crystal. A commonly known and delicate system is chocolate, exhibiting polymorphism with at least six different forms.² Only one of them has the desired texture and taste, but other polymorphs arise from melting and solidification under ambient condition. Another example is the anti-HIV drug ritonavir. It was found to transform to a different polymorphic structure, which is less soluble and, hence, less available.³ All chemical or physical properties accessible from experiment may be heavily changed in the solid state, or generally be sensitive to molecular geometry and environment and, hence, the crystal packing. Examples are NMR chemical shifts,⁴ internal molecular vibrations,^{5,6} or UV/vis excitation and emission behaviour.⁷ Moreover, macroscopic properties like conductivity,⁸ or mechanical properties⁹ naturally arise from the intermolecular interactions in a crystal, not from the intramolecular interactions. But even microscopic changes and their underlying causes may be of interest. The crystal structure of a piano-stool complex (Figure 1.1) shows an unexpected arrangement of the cyclopentadienide protons towards the cation which cannot be reproduced in gas phase calculations. A simple point charge representation of the neighbour molecules, however, could show that interactions with the environment induce the observed structural change.¹⁰

Depending on the desired accuracy it is, therefore, of crucial importance to describe the interactions, be it at a computationally cheap level to rather obtain qualitative trends, or at a very sophisticated level to allow for the correct description of energy differences on a sub-kJ · mol⁻¹ scale.

Unfortunately, calculations become expensive with additional potentials arising and possibly more particles to be treated quantum mechanically. Both liquid and solid environment may be considered, but these are distinctly different in the relative arrangement of molecules. In a liquid, only a local ordering is imposed. For the most part the environment is amorphous and the interactions between molecules are effectively averaged. Therefore, it is possible to describe the solvent as a polarisable continuum¹¹ or sample environment structures. On the other hand, a molecular crystal is very well ordered, thus, providing a very distinct environment. This provides a different challenge, as interactions are not effectively averaged. For a sound description, crystal symmetry must be considered, especially the translational symmetry.

The computationally most accessible way is the use of molecular mechanics (MM) calculations. These allow

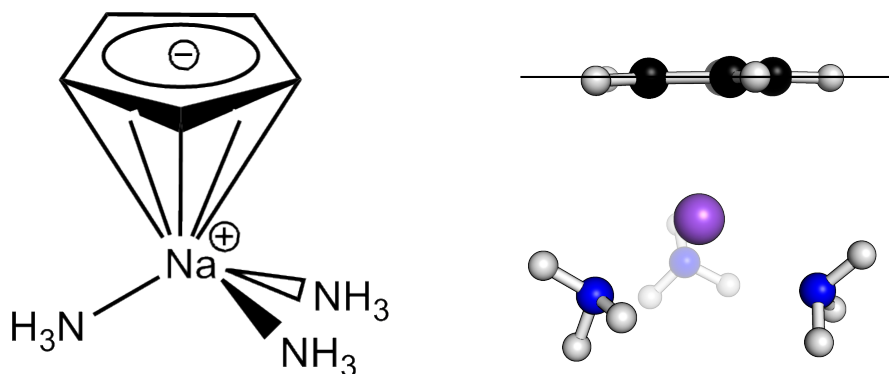


Figure 1.1: Lewis structure and experimental geometry of a piano-stool complex, providing an unusual inclination of the hydrogen atoms slightly towards the cation. This system is studied in detail in Section 5.2.1.

for the calculation of systems with up to millions of atoms, whereas pure quantum mechanical (QM) methods are accessible only up to hundreds or thousands of atoms. Although MM approaches cannot provide accurate energies, they are highly valuable in crystal structure prediction. The variety of space groups combined with different conformations of the molecule and its possible positions in a unit cell leads to a large number of potential polymorphic structures. These can hardly all be validated with QM methods. MM based methods allow for a successful screening of all these structures and help to identify a largely reduced set of promising candidate structures. These candidates can then be assessed with more accurate methods (mostly, but not solely QM approaches) to explore the energy landscape. Such techniques have led to increasing success in crystal structure prediction. This development is observed in the blind test challenges^{12,13,14,15,16,17} held by the Cambridge Crystallographic Data Centre. Despite the increasing difficulty since the beginning, including less rigid systems, a co-crystal and a salt in the sixth test, in three of five cases the experimental structure could be generated and ranked as the global minimum by at least one method. Nevertheless, up to now no method could reliably establish an accurate energy ranking for the promising candidate structures after the MM screening.

On the other hand, there are also success stories when it comes to crystal stability. The use of sophisticated methods could lead to the calculation of the benzene crystal energy within great accuracy (less than $1 \text{ kJ} \cdot \text{mol}^{-1}$) based on the known experimental crystal structure.¹⁸ However, the respective approach relied on a known crystal structure and could not provide an optimisation of the molecule or the cell parameters. Hence, for any practical application one needs to stay somewhere between these extrema of efficiency and accuracy.

Although MM approaches sometimes allow for astonishingly accurate results in the aforementioned blind tests, in general one needs to rely on QM methods or QM/MM hybrid approaches to improve the description. The most straightforward way of including interactions with neighbour molecules, including these molecules in the QM calculation, is unfortunately not viable due to the scaling of computational cost with the system size. Even a finite, tiny crystal of $5 \times 5 \times 5 = 125$ unit cells would contain thousands to ten thousands of atoms. This is still not generally feasible to calculate with pure QM methods.

Therefore, two basic strategies are widely used to reduce the problem from infinite complexity to a feasible size. Periodic wave-function calculations aim to solve the problem with a periodic Hamiltonian, thus, leading to a solution representing the periodicity.¹⁹ For some QM methods efficient formulations exist, but a systematic improvement of the respective methods is often not possible. Moreover, the required solution of the Schrödinger equation in the periodic and non-periodic cases is done somewhat differently. Therefore, techniques developed for the one case are not conveniently transferable to the other. Hence, some well-established gas-phase techniques are essentially infeasible for periodic calculations.

On the other hand, one can follow a *divide et impera* like strategy and deal with individual interactions. This class is very appealing, as it allows for the direct usage of all techniques available for gas phase

calculations. Formally one can even approximate the whole problem with a series of gas phase calculations only, but the convergence of such approaches may be slow. Since the individual contributions require the calculation of two, three, . . . molecules at the same time, both the number of calculations and cost of each calculation quickly increases, limiting its usefulness. Still, it is possible to obtain highly accurate energies, as it was shown in the aforementioned case of the benzene crystal.¹⁸ If less accurate results are acceptable, these effects may as well be included implicitly via approximate potentials, hence, drastically reducing the number of individual calculations required.^{20,21,22} Unfortunately, these methods cannot reasonably reproduce some features arising from the periodic nature of a molecular crystal, e.g. they naturally yield discrete electronic energy levels instead of a continuum electronic band structure.

Finally, one can couple both approaches and e.g. combine an affordable calculation with a periodic Hamiltonian and some molecular energies and intermolecular interaction energies from gas phase calculations. While this promises to merge the best of periodic and non-periodic sides, it still cannot address inherently periodic features, e.g. the band structure, with non-periodic corrections. Nevertheless, such hybrid approaches can quite successfully provide energies or molecular geometries.^{23,24}

This work aims to provide a simple, computationally cheap, and convenient method, as general as gas phase techniques. From the discussion above it is clear that a hybrid QM/MM approach without explicit consideration of a periodic Hamiltonian is the most appropriate way. In contrast to other existing QM/MM approaches, three key points are improved in the context of this work. A common limitation of both QM/MM and periodic calculations is the reference to a unit cell as the building block of a crystal. Although in the former case most methods refer to single molecules and molecule pairs to calculate internal energies and interaction energies, the same set of calculations is required for all molecules in the cell. However, the cell is built from the asymmetric unit, hence, it should be sufficient to calculate this unit to describe the whole. In case the asymmetric unit represents only a part of a molecule, it is favourable to consider a single molecule as the smallest building block. One can reduce the computational cost approximately by the number of molecules per unit cell if one makes use of the fact that molecules are equivalent.

A second point addressed in this work is the unrestricted optimisation of both internal geometry and cell parameters. Common approaches often freeze the cell parameters and, thus, restrict optimisation to the internal geometry. This work not only aims to allow for the optimisation of both, conveniently using readily available atomic gradients, but also makes general use of analytical gradients. In contrast to other approaches requiring the calculation of numerical gradients, all gradient expressions of the presented method are fully analytical. It must, unfortunately, be noted that it relies on QM and MM gradients calculated by the respective external programmes. If these allow for the calculation of analytical gradients for all required QM contributions, one can gain much in computational efficiency, but otherwise the calculation of numerical gradients is the only valid option.

The last guiding principle of the method is a convenient implementation of the different parts (the QM calculation of the one-body energy, the QM calculation of two-body contributions, the calculation of embedding charges, the calculation of MM contributions, the external optimiser). Thus, it makes use of the ChemShell²⁵ QM/MM suite²⁶ which is designed for the convenient coupling of different QM and MM codes. This idea is adopted for the whole model, hence, allowing for the flexible variation of QM theory (individually for charges, one-body energy, and two-body corrections), MM theory, and the optimiser. Individual parts are not required to be implemented within the model, but are handled by external programmes. Therefore, only the interface to such programme needs to be modified (or, possibly, created). Consequently, it becomes possible to use a newly developed method from any QM package in the existing framework at the maximum cost of some tens lines of code to adapt the respective interface file. In contrast to other approaches which are often implemented in the code of an existing QM programme, this provides the user with great flexibility in its usage.

Although the method presented in this work is designed for unrestricted optimisation of molecular crystal structures, it is not targeted at crystal structure prediction. It is based on a given symmetry, and could, hence, only allow for a very limited access to different polymorphs by optimisation. While symmetry is accounted

for to reduce the computational cost, no explicit and extensive use is made of the full crystal symmetry, nor is the actual symmetry enforced. Thorough evaluations of different theory parameters are done to establish a reasonable model, but the whole scheme is built within the framework of existing potentials. This work does not feature the development and evaluation of new or modified potentials to describe intermolecular interactions. Finally, it should be noted that the presented method aims to be computationally feasible and convenient. Albeit results can be systematically converged to high accuracy, the key target is to improve upon a simple gas description.

2 Theory

2.1 Wave-function Theory

One fundamental axiom of quantum mechanics is the existence of a wave-function Ψ which implicitly contains all information on the system. The application of a proper operator \hat{O} extracts the corresponding physical observable O as the eigenvalue according to the equation

$$\hat{O}\Psi = O\Psi. \quad (2.1)$$

A fundamental equation for quantum mechanical applications is the Schrödinger equation

$$i\hbar \frac{\partial \Psi(t)}{\partial t} = \hat{H}(t)\Psi(t), \quad (2.2)$$

which connects the time-dependent change of a wave-function with the change of the Hamiltonian operator $\hat{H}(t)$ over time. If the Hamiltonian is time-independent, the system is in a static state. The time-dependent Schrödinger equation can be reduced to the time-independent formulation²⁷

$$\hat{H}\Psi = E\Psi, \quad (2.3)$$

where E represents the energy of the state.

The Hamiltonian for a system of N_n positively charged nuclei (A,B,...) and N_e negatively charged electrons (a,b,...) at least contains two types of contributions, Coulomb interactions V among the particles and kinetic energy T of each particle. In the absence of an external field, the Hamiltonian is

$$\hat{H} = \sum_{A,B}^{N_n} \frac{Z_A Z_B}{r_{AB}} - \sum_A^{N_n} \sum_a^{N_e} \frac{Z_A}{r_{Aa}} - \sum_{a,b}^{N_e} \frac{1}{r_{ab}} - \frac{1}{2} \sum_A^{N_n} \frac{1}{M_A} \nabla_A^2 - \frac{1}{2} \sum_a^{N_e} \nabla_a^2. \quad (2.4)$$

The first three terms describe the Coulomb interaction among nuclei with nuclear charges Z , between nuclei and electrons, and among electrons. The latter terms represent the kinetic energy of nuclei and electrons, respectively. An often very useful approximation was introduced by Born and Oppenheimer.²⁸ Since electrons are by at least three orders of magnitude lighter than nuclei, they experience much larger acceleration and velocity. Hence, it becomes possible to separate the electronic and nuclear movement and solve the respective equations of motion separately. The problem then decomposes into an electronic and nuclear Schrödinger equation, which can be solved subsequently. The electronic Schrödinger equation

$$\hat{H}_e \Psi_e(\vec{R}) = E(\vec{R}) \Psi_e(\vec{R}) \quad (2.5)$$

with the electronic wave-function Ψ_e and Hamiltonian

$$\hat{H}_e = \sum_{A,B}^{N_n} \frac{Z_A Z_B}{r_{AB}} - \sum_A^{N_n} \sum_a^{N_e} \frac{Z_A}{r_{Aa}} + \sum_{a,b}^{N_e} \frac{1}{r_{ab}} - \frac{1}{2} \sum_a^{N_e} \nabla_a^2 \quad (2.6)$$

still contains all nuclear coordinates \vec{R} , but these are now constants, not variables. Consequently, the electronic Hamiltonian does not contain the nuclear kinetic energy, and the internuclear Coulomb energy

is a constant value for the respective nuclear geometry. Then Equation 2.5 can be solved for an arbitrary nuclear configuration, and the resulting energies may be used to build a potential energy surface to solve the nuclear Schrödinger equation. Albeit the complexity of the problem is reduced with this approximation, it remains (except for one electron systems) a many-body problem. It can only be solved analytically for the simplest cases (e.g. the hydrogen atom). Therefore, numerical approaches are applied to obtain an approximate solution. Since all of the methods applied in this work (and presented in the following sections) are considered with the approximate solution of the electronic Schrödinger equation, the respective electronic wave-function is in the following shortly denoted as Ψ , not Ψ_e . Similarly, the electronic Hamiltonian \hat{H}_e is always denoted as \hat{H} in all following sections.

2.1.1 The Hartree-Fock Method

A series of techniques is based on the approximate calculation of a wave-function, hence, denoted as wave-function theory (WFT). The foundation of most wave-function based descriptions is the Hartree-Fock method.²⁹ In order to obey the Pauli principle,³⁰ the wave-function is conveniently represented with a Slater determinant³¹

$$\psi_{\text{SD}} = \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_{N_e}(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_{N_e}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N_e) & \phi_2(N_e) & \cdots & \phi_{N_e}(N_e) \end{vmatrix}. \quad (2.7)$$

It contains a set of one-electron orbitals which need to be determined.

For convenience, the Dirac bracket notation is used, whereby

$$\begin{aligned} \int \phi_i^* \phi_j d\tau &= \langle i|j \rangle \\ \int \phi_i^* \hat{O} \phi_j d\tau &= \langle i|\hat{O}|j \rangle \end{aligned} \quad (2.8)$$

with arbitrary wave-functions ϕ and the electron coordinates (spatial and spin) represented with τ .

The Hamiltonian (Equation 2.6) can be separated into three parts. As it does not depend on the electronic wave-function, the internuclear interaction can be dropped for the solution of the Schrödinger equation. Considering the other terms, electronic kinetic energy and electron-nucleus interactions depend on one electron coordinate, while the electron-electron interaction depends on two electron coordinates. Hence, the Hamiltonian can be written as the sum of one-electron operators (\hat{h}_a) and two-electron operators ($\frac{1}{r_{ab}}$)

$$\hat{H} = \sum_a \hat{h}_a + \sum_{a<b} \frac{1}{r_{ab}}. \quad (2.9)$$

The integration of the Schrödinger equation over the complete space yields an energy expression

$$E = \langle \Psi|\hat{H}|\Psi \rangle. \quad (2.10)$$

The use of a Slater determinant (Equation 2.7) for Ψ and the separated Hamiltonian (Equation 2.9) yields a large number of terms. Fortunately, if the ϕ are required to be orthonormal, only elements of the form

$$\begin{aligned} \langle \phi_i(1)|\hat{h}_i(1)|\phi_i(1) \rangle &\equiv (i|\hat{h}|i) \\ \langle \phi_i(1)\phi_j(2)|\frac{1}{r_{12}}|\phi_i(1)\phi_j(2) \rangle &\equiv (ii|jj) \\ \langle \phi_i(1)\phi_j(2)|\frac{1}{r_{12}}|\phi_j(1)\phi_i(2) \rangle &\equiv (ij|ji) \end{aligned} \quad (2.11)$$

remain non-zero.

An effective one-electron operator, the Fock operator

$$\hat{F}_i = \hat{h}_i + \sum_j \left(\hat{J}_j - \hat{K}_j \right), \quad (2.12)$$

can be constructed, fulfilling the pseudo-eigenvalue equations

$$\hat{F}_i \phi_i = \epsilon_i \phi_i \quad (2.13)$$

with the orbital energy ϵ_i . \hat{J} and \hat{K} generate two-electron integrals with the residual orbitals ϕ_j . Therefore, a change in any orbital ϕ_j affects the Fock operator \hat{F}_i and Equations 2.13 need to be solved iteratively. The resulting Slater determinant of the HF procedure serves as the starting point for other methods. It is denoted as Φ_0 .

The Hartree-Fock energy is obtained as

$$E_{\text{HF}} = \sum_{i=1}^{N_e} \left(i | \hat{h} | i \right) + \frac{1}{2} \sum_{ij}^{N_e} \left((ii|jj) - (ij|ji) \right). \quad (2.14)$$

Except for one-electron systems, this is not the exact electronic energy obtainable for any given one-particle basis set. The respective difference to the exact energy is referred to as the correlation energy. Different methods have been developed to approximate the correlation energy. These electron correlation calculations with an arbitrary method M generate an energy correction ΔE_M to be added to the HF energy to give the correlated energy

$$E_M = E_{\text{HF}} + \Delta E_M. \quad (2.15)$$

2.1.2 Møller-Plesset Perturbation Theory

One approach to improve on the HF solution is Møller-Plesset perturbation theory (MP).³² The foundation of a perturbational treatment is that an approximate solution to a problem is known and the difference may be described with a small perturbation. One uses

$$\Psi = \lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \dots \quad (2.16)$$

for the wave-function and

$$\hat{H} = \lambda^0 \hat{H}_0 + \lambda^1 \hat{H}_1 \quad (2.17)$$

for the Hamiltonian. The respective Ψ_n and \hat{H}_n are the n th-order wave-function and Hamiltonian, $\lambda \in [0, 1]$ varies the strength of the perturbation. After these are combined and the result is reordered, a series of expressions for the n th-order energy corrections

$$W_n = \left\langle \Psi_0 | \hat{H}_1 | \Psi_{n-1} \right\rangle \quad (2.18)$$

is obtained. This can be recursively combined, so, one can find that the n th-order wave-function is needed to calculate the $(2n + 1)$ th-order energy correction. Hence, until MP3 only the first-order wave-function is required. From a HF calculation a set of occupied and virtual (unoccupied) orbitals is obtained. In Φ_0 only the occupied orbitals are contained. However, an electron may be excited from an occupied orbital i to a virtual orbital a , hence, generating an excited configuration Φ_i^a . In the same way configurations with up to all N_e electrons are possible, leading to Φ_{ij}^{ab} , Φ_{ijk}^{abc} , and so on. A linear combination

$$\Psi_1 = \sum_i c_i \Phi_i \quad (2.19)$$

of all excited determinants Φ_i may be used to approximate the first-order wave-function. Hereof a closed expression for the coefficients c_i can be derived, finally leading to an expression for the second-order energy correction

$$W_2 = \sum_{i \neq 0} \frac{\langle \Phi_0 | \hat{H}_1 | \Phi_i \rangle \langle \Phi_i | \hat{H}_1 | \Phi_0 \rangle}{E_0 - E_i}. \quad (2.20)$$

In the special case of Møller-Plesset perturbation theory, the \hat{H}_0 is chosen as the sum over all Fock operators, and \hat{H}_1 as r_{12}^{-1} . Due to this construction, the HF energy is not the zeroth-order, but the first-order energy. Therefore, the first improvement upon HF is provided by the second-order correction. It can be shown that only doubly excited determinants can contribute to elements of the form $\langle \Phi_i | \hat{H}_1 | \Phi_0 \rangle$, therefore, only these need to be considered. Making further use of the orthogonality between the orbitals ϕ and Brillouin's theorem, the final expression

$$\Delta E_{\text{MP2}} = \sum_{i < j} \sum_{a < b} \frac{((ia|jb) - (ib|ja))^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (2.21)$$

for the MP2 energy correction can be derived.

Spin-Component Scaling

The formulation of the MP2 energy in Equation 2.21 does not consider spin explicitly. All contributions are treated equally. However, one may split it into terms arising from orbitals of the same spin (ss) and opposite spin (os), which are then weighted differently, referred to as spin-component scaling (SCS).³³ In the SCS-MP2 scheme scaling parameters $p_{\text{ss}} = 1/3$ and $p_{\text{os}} = 6/5$ are introduced, leading to a modified expression of the MP2 energy correction

$$\Delta E_{\text{MP2}} = p_{\text{ss}} \Delta E_{\text{ss}} + p_{\text{os}} \Delta E_{\text{os}} \quad (2.22)$$

with the spin components

$$\begin{aligned} \Delta E_{\text{ss}} &= \sum_{i < j} \sum_{a < b} \frac{((ia|jb) - (ib|ja))^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} + \sum_{\bar{i} < \bar{j}} \sum_{\bar{a} < \bar{b}} \frac{((\bar{i}\bar{a}|\bar{j}\bar{b}) - (\bar{j}\bar{b}|\bar{i}\bar{a}))^2}{\epsilon_{\bar{i}} + \epsilon_{\bar{j}} - \epsilon_{\bar{a}} - \epsilon_{\bar{b}}} \\ \Delta E_{\text{os}} &= \sum_{i\bar{j}} \sum_{a\bar{b}} \frac{((ia|\bar{j}\bar{b}))^2}{\epsilon_i + \epsilon_{\bar{j}} - \epsilon_a - \epsilon_{\bar{b}}}. \end{aligned} \quad (2.23)$$

The over-lined orbitals represent β spin orbitals. Scaling factors are not derived from first principles, but optimised to improve agreement of SCS results with experimental data.³³

2.1.3 Coupled-Cluster Theory

The central idea of coupled cluster theory,³⁴ as it is introduced in theoretical chemistry,³⁵ is to use the electron excitation operator \hat{T} not directly, but in an exponential form to generate the coupled cluster wave-function

$$\Psi_{\text{CC}} = e^{\hat{T}} \Phi_0. \quad (2.24)$$

\hat{T} may include all operators for the excitation of one to N_e electrons, but is usually truncated. In the case of CCSD \hat{T} is restricted to the single and double excitation operators \hat{T}_1 and \hat{T}_2 . Then a Taylor expansion of the exponential yields an infinite series of terms,

$$e^{\hat{T}} = e^{\hat{T}_1 + \hat{T}_2} = 1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}(\hat{T}_1 + \hat{T}_2)^2 + \dots = 1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}(\hat{T}_1^2 + 2\hat{T}_1\hat{T}_2 + \hat{T}_2^2) + \dots, \quad (2.25)$$

which includes higher-order excitations through the coupling of \hat{T}_1 and \hat{T}_2 . These act on the HF solution to generate all singly and doubly excited configurations

$$\begin{aligned}\hat{T}_1\Phi_0 &= \sum_i^{\text{occ}} \sum_a^{\text{virt}} t_i^a \Phi_i^a \\ \hat{T}_2\Phi_0 &= \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} t_{ij}^{ab} \Phi_{ij}^{ab}\end{aligned}\tag{2.26}$$

with the amplitudes t determining the weight of the respective configuration. It is not possible to determine these directly, unlike the coefficients in MP2, but a series of coupled equations can be derived which allows for an iterative determination.

From the Schrödinger equation 2.5 with the expansion of $e^{\hat{T}}$ (Equation 2.25) and the operator definitions (2.26) plugged in, a projection onto the reference wave-function yields a coupled cluster energy expression

$$E_{CC} = \left\langle \Phi_0 | \hat{H} (\hat{T}_2 + \frac{1}{2} \hat{T}_1^2) | \Phi_0 \right\rangle.\tag{2.27}$$

Due to Brillouin's theorem, singly excited configurations do not contribute, and since \hat{H} only contains one- and two-electron operators, triply and higher excited determinants do not contribute as well. Hence, a coupled cluster energy correction

$$\Delta E_{CC} = \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) ((ia|jb) - (ib|ja))\tag{2.28}$$

can be calculated.

In the CCSD(T) scheme,³⁶ an additional contribution (T) to the correlation energy is considered. It mixes the coupled cluster and MP approaches by using energy expressions from MP4 with the coupled cluster amplitudes used as coefficients.

2.1.4 Local Electron Correlation Methods

The orbitals one obtains from a HF calculation are usually delocalised over the whole molecule. For the application of electron correlation methods this represents an undesirable feature, since essentially all molecular orbitals have a significant overlap. The resulting contributions of integrals $(ia|jb)$ with arbitrary virtual orbitals $\phi_{a,b}$ and occupied orbitals $\phi_{i,j}$ are, therefore, non-negligible. It results in a steep increase of the computational cost with the system size. Due to this unfortunate scaling, much effort has been invested in the application of local orbital based correlation methods. Ideally, one would be able to localise orbitals in a way that only requires a somewhat constant number of excitations between spatially close localised orbitals to be calculated. If these cover the majority of the correlation energy, other excitations can be neglected and a near-linear scaling might be achieved.

In general, a new set of orbitals $\tilde{\phi}_j$ can be generated by an arbitrary unitary transformation of the HF orbitals according to

$$\tilde{\phi}_j = \sum_i U_{ij} \phi_j.\tag{2.29}$$

In order to obtain local orbitals from this transformation, one needs to apply constraints which in some way consider the locality of the transformed orbitals. The Pipek-Mezey scheme³⁷ employs a maximum occupation at preferably few atoms sites as the localisation criterion. For each occupied orbital a set of virtual orbitals to which excitations are allowed is selected, referred to as a domain. The selection is made according to natural population analysis (NPA)³⁸ occupation numbers. These are known to be very stable with different basis sets, resulting in often basis set independent domains.³⁹

A secondary advantage of local correlation methods is the possible reduction of the basis set superposition error (BSSE) if the virtual space consists of projected atomic orbitals (PAOs).⁴⁰ Due to the incompleteness of

a finite basis set, the presence of additional basis functions on the second molecule during a dimer calculation in principle allows for a more complete description of the orbitals. Hence, the energy of a molecule dimer is to some extent artificially lowered, as additional basis functions from the second monomer are available compared to the respective monomer calculations. Interaction energies from a supramolecular approach, therefore, systematically overestimate the stabilisation of a dimer, albeit the effect becomes less important the larger the basis set is chosen. In the PAO formalism, local orbitals by construction enforce the use of basis functions only at few atomic centers. Hence, the influence of basis functions from a second monomer is greatly reduced. Consequently, dimer calculations with localised orbitals suffer much less from the intermolecular BSSE and can potentially provide more reliable interaction energies.

2.2 Density Functional Theory

Wave-function based methods always consider electrons individually. Therefore, the wave-function will depend on $4N_e$ coordinates. In contrast, density functional theory (DFT) aims to describe a system with all its properties solely on the basis of the electron density.⁴¹ This is a function of only three spatial coordinates. This greatly reduces the complexity of the problem, but unfortunately for the price of new challenges arising. In DFT all electron contributions need to be replaced with analogous density contributions, and the sum over electrons is replaced with an integral over the whole space. However, the exact functional forms for kinetic, exchange and correlation energy of electrons are unknown and can, thus, only be approximated. Hence, in the DFT energy expression

$$E_{\text{DFT}}[\rho] = V_{\text{nn}} + V_{\text{ne}}[\rho] + V_{\text{ee}}[\rho] + T_e[\rho] + E_x[\rho] + E_c[\rho] \quad (2.30)$$

the Coulomb energy terms V can be calculated exactly for a given density ρ , but for the electronic kinetic energy T_e , exchange energy E_x , and correlation energy E_c this is not possible. It was found that the description of the kinetic energy based on the density is very poor, hence, a different approach is required. Great improvement was gained from the introduction of molecular orbitals to DFT in the framework of Kohn-Sham (KS) theory.⁴² A similar expression for the orbitals as in HF theory (Equation 2.13) can be obtained in the KS framework

$$\hat{h}_i^{(\text{KS})} \phi_i = \epsilon_i \phi_i \quad (2.31)$$

with the Kohn-Sham operator $\hat{h}_i^{(\text{KS})}$. It contains the same formulation of kinetic energy and Coulomb energy as the Fock operator, but unlike the latter it has exchange and correlation potentials from the respective functional description. The orbital description greatly improves the kinetic energy estimate, although the introduction of orbitals increases the dimensionality from 3 to $4N_e$. Nevertheless, the description of exchange and correlation is generally not based on orbitals. Hence, the complexity of DFT even with orbitals is much reduced in comparison to WFT. On the other hand, the form of the exact exchange and correlation functionals are not known. Therefore, a systematic improvement, unlike in WFT, is not possible in DFT. Still, the formal scaling of KS DFT with the number of basis functions is only N_b^3 , if density fitting techniques are used. This allows for a broad applicability of DFT to large molecules, which is not generally the case for more accurate WFT methods.

2.2.1 Ranks of Density Functional Theory

The simplest assumption possible for the exchange and correlation energies is a locally uniform density, i.e. at every point in space the density is assumed to be approximately uniform. In that case formulae which are exact for a true uniform electron gas are applied. Alas, the approximation of a slowly varying density does not describe molecular systems well and leads to significant errors. In order to account for the variation, one may not only use the density, but as well the gradient of the density $\nabla\rho$. Then the exchange and correlation energies become functionals of both ρ and $\nabla\rho$, which is referred to as the generalised gradient approximation

(GGA). In principle one needs not limit the description to the first derivative, but can as well make use of the second (and third, ...) derivative. Density functionals making use of higher order derivatives are referred to as meta-GGA.

Since orbitals have been introduced with KS, the exact exchange energy may be evaluated like in HF. Hence, for the KS orbitals the terms from an exchange functional and the HF exchange K may be mixed to yield

$$E_x = \alpha E_x[\rho] + (1 - \alpha)K(\phi) \quad (2.32)$$

with a scaling factor α . This approach is usually known as hybrid DFT. The calculation of HF exchange increases the computational cost to the same level as HF. However, in contrast to HF it offers an improved description, because electron correlation is already considered.

2.2.2 Dispersion Corrections

One well known flaw of basic density functional theory is the inability to describe dispersion interactions. It has been attributed to the local form of standard functionals calculating exchange and correlation contributions only based on the density and its gradients at points in space, but independent of the surrounding density. One approach to address this problem is the introduction of parametrised correction terms for DFT in the framework of the D3 model.⁴³ In general this model makes use of well-known parametric forms of dispersion contributions. Two-body terms $E^{(2)}$ and three-body terms $E^{(3)}$ are considered to calculate the dispersion energy. The former one is given as

$$E_{\text{disp}}^{(2)} = \sum_{AB} \sum_{n=6,8} s_n \frac{C_n^{AB}}{r_{AB}^n} f_{d,n}(r_{AB}), \quad (2.33)$$

where all pairs of atoms A and B are summed up. There s_n is a scaling factor, the C_n^{AB} represent the dispersion coefficients of order n , r_{AB} is the atom distance, and $f_{d,n}$ is a damping function. The latter is required, since the short range contributions of dispersion are captured with local DFT. Hence, a double-counting of these contributions is avoided with damping of the correction term for short distances. The C_6 dispersion coefficients are derived from time-dependent DFT according to the Casimir-Polder formula, higher order terms are determined recursively from C_6 . Within the D3 model only the C_6 and C_8 terms are used.

With increasing system size, the importance of three-body terms $E^{(3)}$ to the total dispersion energy grows as well. For this contribution, all triples of atoms A , B , and C need to be considered as

$$E^{(3)} = \sum_{ABC} \frac{C_9^{ABC} (3 \cos \theta_a \cos \theta_b \cos \theta_c + 1)}{(r_{AB} r_{AC} r_{BC})^3}. \quad (2.34)$$

Here θ_a , θ_b , and θ_c represent the angles in the ABC triangle, and r_{AB} , r_{AC} , and r_{BC} are the pairwise atom distances. The C_9 coefficients are in principle defined from time-dependent DFT, but in the D3 model an approximation based on the C_6 constants is made.

2.3 Symmetry-Adapted Perturbation Theory

Symmetry-adapted perturbation theory (SAPT)⁴⁴ is a perturbation theory approach to the calculation of intermolecular interaction energies. Within the SAPT framework, single molecule calculations are carried out to build the unperturbed system. Then the interaction between two monomers A and B is calculated from an anti-symmetrised product of the monomer wave functions Ψ_A and Ψ_B with a perturbed Hamiltonian \hat{H} . The latter includes the unperturbed Hamiltonians \hat{H}_A and \hat{H}_B , and an interaction operator \hat{V} which represents the perturbation. A series of interaction terms arises. If terms up to the second order are included,

the interaction energy

$$\Delta E(AB) = \Delta E_{\text{elst}}^{(1)} + \Delta E_{\text{exch}}^{(1)} + \Delta E_{\text{ind}}^{(2)} + \Delta E_{\text{disp}}^{(2)} + \Delta E_{\text{exch-ind}}^{(2)} + \Delta E_{\text{exch-disp}}^{(2)} \quad (2.35)$$

consists of the contributions of electrostatic interaction ($\Delta E_{\text{elst}}^{(1)}$), polarisation ($\Delta E_{\text{ind}}^{(2)}$), and dispersion ($\Delta E_{\text{disp}}^{(2)}$), all arising from the Coulomb interaction between the systems. Complementary, the exchange contribution leads to the terms considered as first-order exchange ($\Delta E_{\text{exch}}^{(1)}$), exchange-induction ($\Delta E_{\text{exch-ind}}^{(2)}$), and exchange-dispersion ($\Delta E_{\text{exch-disp}}^{(2)}$).

Although in principle intermolecular correlation can be described together with the intermolecular interactions (leading to a double perturbation approach), the description becomes much more difficult. Hence, standard SAPT is based on HF monomer wave-functions (HF-SAPT), and intramolecular correlation is neglected. However, if the wave-function for SAPT is taken from KS orbitals (DFT-SAPT), correlation is implicitly accounted for.⁴⁵ Therefore, a perturbation ansatz for the intramolecular correlation is not required. As the monomer densities improve compared to the HF density, the description of the interaction between them improves as well. Electrostatic terms could even become exact (for the exact density), while the exchange contributions still depend on the accuracy of the exchange functional. Nonetheless, DFT-SAPT provides much more accurate interaction energies than HF-SAPT and can compete with supramolecular approaches making use of high-level electron correlation methods.⁴⁶ Moreover, the SAPT procedure is by construction free of the intermolecular BSSE, as the monomer wave functions are used to calculate the interaction.

2.4 General Hybrid QM/MM

Unlike quantum mechanics, molecular mechanics makes use of fitted analytical functions for the energy. In the MM frame the energy of a molecular system is commonly calculated from different contributions, both bonding (e.g. bond stretching and angle bending) and non-bonding (e.g. electrostatics). While all terms need to be included in a pure MM calculation, a hybrid QM/MM description may neglect some of them under certain circumstances. In case the MM part does not contribute its internal energy, all interactions among MM atoms can simply be neglected. If QM and MM parts are not connected with chemical bonds, only non-bonded potentials are needed to describe the interaction of the MM region with the QM region. If the two aforementioned conditions are met, no bonded interactions are required at all, and mainly two contributions remain. Coulomb interactions are in the simplest case represented with atomic point charges at the MM atoms. Their interaction with the QM part is very simple. The N_q point charges q are treated according to the Coulomb law. One additional term is added to the pure QM Hamiltonian \hat{H}_{QM} to yield the Hamiltonian in the point charge (PC) field

$$\hat{H}_{\text{QM+PC}} = \hat{H}_{\text{QM}} + \sum_i^{N_e} \sum_j^{N_q} \frac{q_j}{r_{ij}} \quad (2.36)$$

The additional term can be treated exactly like the electron-nucleus interactions. Consequently, the respective interactions are already included in the final QM energy and do not need to be considered explicitly.

Even simpler is the inclusion of van-der-Waals (vdW) terms in a QM/MM description. The approximation is made that the latter do not influence the wave-function. Hence, they do not enter the Hamiltonian at all. Instead, they are described with pairwise potentials, often in the form of 12-6 Lennard-Jones (LJ) potentials

$$V_{\text{LJ}} = \epsilon \left(\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 \right) \quad (2.37)$$

describing the potential energy V at a distance r . The equilibrium distance r_0 and the well depth ϵ are parameters dependent on the two atom types the interaction takes place between. These terms are summed

up for all pairs of one QM and one MM atom, as the internal MM contributions are neglected. The QM calculation is not affected by these contributions. Only a contribution to energies and gradients needs to be included in the construction of QM/MM energies and gradients. For a frozen MM environment, this yields the final energy expression

$$E_{\text{QM/MM}} = E_{\text{QM+PC}} + E_{\text{LJ}} \quad (2.38)$$

with the QM energy including the electrostatic interaction contribution.

2.5 Electronic Structure Calculations of Molecular Crystals

While the description of a finite number of atoms is straightforward, the steep scaling of QM based methods with the system size limits their applicability. Fortunately, crystals are characterised by the translational symmetry of a unit cell, hence, the whole system can be seen as multiple identical copies of the same unit. Making use of the translational symmetry makes the treatment of such systems amenable. There are two general approaches to make use of the crystal symmetry. The first way is the explicit consideration of wave-function periodicity, the second way is to break the crystal down to smaller parts that can be treated individually.

2.5.1 Periodic Wave-function Approaches

In a periodic system all cells are equivalent. Therefore, the Hamiltonian and wave-function must both be invariant to translation by a vector

$$\vec{T} = m\vec{a} + n\vec{b} + o\vec{c} \quad (2.39)$$

with integer numbers m, n, o and real-space cell vectors $\vec{a}, \vec{b}, \vec{c}$. So, it is

$$\hat{H}\Psi(\vec{r}) = E\Psi(\vec{r}) \quad (2.40)$$

with $\Psi(\vec{r}) = \Psi(\vec{r} + \vec{T})$. Solutions to this equation are Bloch functions.¹⁹ These must fulfil the condition

$$\Phi(\vec{r} + \vec{T}, \vec{k}) = \exp(i\vec{k}\vec{T})\phi(\vec{r}, \vec{k}) \quad (2.41)$$

with \vec{k} being a wave-vector. It is found that \vec{k} must be composed of integer multiples of the inverse cell vectors, which corresponds to an infinite number of choices. However, one has to restrict calculations to a limited number of \vec{k} -vectors for practical reasons. To fulfil Equation 2.41, functions of the form

$$\Phi(\vec{r}, \vec{k}) = \exp(i\vec{k}\vec{r})\Omega(\vec{r}, \vec{k}), \quad (2.42)$$

are required. The exponential term represents a plane-wave, and the function Ω must represent the periodicity of the lattice. Comparably to non-periodic calculations, the crystal orbitals are expanded in basis functions. If atom centered basis functions χ are used, all translationally equivalent atoms have to be included to form a periodic function

$$\Omega = N^{-1/2} \sum_{\vec{T}} \exp(i\vec{k}(\vec{T} - \vec{r}))\chi_{\vec{T}}(\vec{r} - \vec{T}). \quad (2.43)$$

The normalisation factor $N^{-1/2}$ depends on the number N of unit cells considered, the summation covers atomic orbitals for all translation vectors \vec{T} . Hence, a formulation of the functions

$$\Phi(\vec{r}, \vec{k}) = N^{-1/2} \sum_{\vec{T}} \exp(i\vec{k}\vec{T})\chi(\vec{r} - \vec{T}) \quad (2.44)$$

arises. Crystal orbitals are then composed from N_b Bloch functions as

$$\phi_i(\vec{r}, \vec{k}) = \sum_{\alpha}^{N_b} c_{i\alpha}(\vec{k}) \Phi_{\alpha}(\vec{k}). \quad (2.45)$$

An analogous formulation of the HF Equation 2.13 can be obtained with these orbitals. Since they depend on the wave-vector, a whole set

$$\hat{F}_i(\vec{k})\phi_i(\vec{k}) = \epsilon_i(\vec{k})\phi_i(\vec{k}) \quad (2.46)$$

of equations results. These need to be solved for every \vec{k} chosen. Hence, not a set of discrete energy states, but a continuum of states results, commonly known as the band structure of a solid.

The Coulomb part of periodic calculations is more challenging than for finite systems, as the interaction only slowly decays and simple cutoffs are not sufficient. Fortunately, an efficient treatment via e.g. Ewald summation circumvents this problem and makes pure DFT calculations feasible. Unfortunately, the exchange part in HF and hybrid DFT calculations is slowly decaying and becomes very expensive in periodic calculations.⁴⁷ Consequently, the applicability of hybrid DFT functionals is still today limited to much smaller unit cells than (meta-)GGA functionals. WF calculations, which require not only a periodic HF, but also a subsequent electron correlation calculation to improve upon DFT, are even more challenging.

Periodic MP2 theory has been applied quite some time ago,⁴⁸ and more recent approaches^{49,50,51} making use of localised orbitals increase its efficiency. Nonetheless, it is still far from being as applicable as standard MP2 for non-periodic systems. Moreover, rather large basis sets are required to obtain reasonably converged correlation energies, but in the case of Gaussian basis sets this may lead to linear dependency problems. Beyond MP2, CCSD(T) can routinely be applied to small to medium sized molecules in the gas phase, while local implementations even allow for the calculation of hundreds of atoms.¹ For periodic systems, however, only very few applications limited to small systems have been reported so far.⁵²

2.5.2 Additive Approaches

A second family of approaches to molecular crystals makes use of a fragmentation, so, the infinite total system is partitioned into finite subsystems. Basically, one assumes that the total energy E can be calculated as

$$E = \sum_A E(A) + \sum_{AB} \Delta E(AB) + \sum_{ABC} \Delta E(ABC) + \dots \quad (2.47)$$

from the energies of non-interacting subsystems A, B, C, \dots and the interaction energies between pairs, triples, \dots of subsystems. If the contributions are added up completely, the description is exact, but one usually truncates the expansion. The question is how to define the expansion to ensure convergence. In the case of molecular crystals, single molecules are the natural choice for subsystems. The contributions $\Delta E(AB)$ and $\Delta E(ABC)$ are obtained from calculations of dimers and trimers, which can be done routinely with common QM methods. Formally, one needs to consider an infinite number of energies $E(A)$ and interactions $\Delta E(AB)$ to obtain the total energy. However, given the translational symmetry one can restrict the sum to the molecules of one unit cell UC (or even a single molecule, if all molecules in the cell are symmetry equivalent). The interaction energies are expected to decay with increasing intermolecular distance. Thus, these may be restricted according to a distance criterion to a limited set of molecules from nearby cells NC . Hence, for a description up to the two-body level a rather general expression of the form

$$E = \sum_{A \in UC} E(A) + \frac{1}{2} \sum_{A \in UC} \sum_{B \in UC} \Delta E(AB) + \sum_{A \in UC} \sum_{B \in NC} \Delta E(AB) \quad (2.48)$$

arises. Correction terms $\Delta E(AB)$ are commonly calculated from a supramolecular approach as

$$\Delta E(AB) = E(AB) - E(A) - E(B) \quad (2.49)$$

with the energies of the dimer AB and the monomers A and B . This approach is unfortunately not generally sufficient for the description of molecular crystals. One is often required to include third (or even higher order) contributions. This strongly limits the applicability of such an approach, as the number of required trimers and computational cost of each trimer calculation increase rather steeply. An approach going up to four-body contributions treated with coupled cluster theory¹⁸ could provide an impressively accurate energy of the benzene crystal from an experimental geometry. One can, however, try to simulate the higher order effects differently.

The Embedded Many-body Expansion

In the embedded many-body expansion (EMBE) of Bygrave, Allan and Manby,²⁰ energy contributions are considered up to the two-body level explicitly, but higher order terms are included implicitly. Potentials for Coulomb and exchange repulsion are included in all monomer and dimer calculations. For the Coulomb potential spherical Gaussian density distributions at atomic positions a of the molecules are fitted to represent the HF density of single monomers. The Coulomb potential

$$v(\vec{r}) = \sum_{a \notin AB} \int \frac{\rho_a(\vec{r}_a)}{|\vec{r} - \vec{r}_a|} d\vec{r}_a \quad (2.50)$$

at position \vec{r} results from the embedding densities ρ . For each set of calculations only potentials of atoms which are not part of the respective molecules are included. The resulting potential is then added to the nuclear Coulomb energy and the one-electron part of the electronic Hamiltonian. Exchange repulsion terms are included with an approximate potential. The exchange-repulsion term

$$E_{\text{rep}} \approx kS_{AB} \quad (2.51)$$

is approximated with a system-specific constant k and the density overlap

$$S_{AB} = \int \rho_A(\vec{r})\rho_B(\vec{r})d\vec{r} \quad (2.52)$$

between the densities of monomers A and B . The constant k is fitted to E_{rep} and S_{AB} from SAPT calculations. A counterpoise correction is applied to eliminate the BSSE, so, corrected monomer energies are calculated as

$$\Delta\Delta E_{AB} = -(E_A(B) - E_A + E_B(A) - E_B) \quad (2.53)$$

with $E_A(B)$ representing the monomer energy of A calculated with basis functions of B present. The approach makes use of symmetry, so, equivalent interactions have to be computed only once. No analytical gradients are available, but massive parallelisation makes optimisation feasible at least for small systems at the MP2 level of theory.

The Crystallographic QM/MM Model

The Crystallographic QM/MM (c-QM/MM) model²¹ is an approach which is methodically very similar to the EMBE model. Two key differences, namely the additional use of crystal symmetry and the simpler description of many-body contributions, have to be noted. In the EMBE framework, a whole unit cell is calculated. Therefore, the sums over all molecules A of the cell in Equation 2.48 arise. The c-QM/MM model makes full use of the crystal symmetry, transforming equation 2.48 to

$$E = E(A) + \sum_{B \in UC} \Delta E(AB) + \sum_{B \in NC} \Delta E(AB). \quad (2.54)$$

This may, as long as all molecules in the cell are symmetry equivalent, reduce the computational cost by the number of molecules per cell. Secondly, the c-QM/MM scheme aims to include higher order effects implicitly,

but in a simpler and more flexible manner. Instead of Gaussian charge distributions, simple point charges at atomic positions are used, and 6-exp parametrised potentials are included to account for vdW interactions. The model allows for the relaxation of cell parameters, but does not include the optimisation of molecular geometries.

Embedded QM/MM

The embedded QM/MM scheme developed by Bjornsson and Bühl²² presents a simple but computationally cheap and flexible model for the consideration of environment effects. The crystal is reduced to a finite cluster, separated into a QM region (one unit cell) and an MM region surrounding the QM part. MM internal contributions are neglected. Only interactions with the QM region are taken into consideration. Two types of interactions are used, electrostatic and vdW interactions. For the electrostatics part, QM derived point charges are placed at the MM atom sites, and additional charges are placed at the cluster surface to mimic the periodic electrostatics. Atomic charges are determined repeatedly from embedded QM calculations to achieve a self-consistent set of charges. The vdW contributions are included with pairwise LJ potentials. A single unit cell is calculated and may be optimised, while the surrounding MM geometry is frozen. The restriction of the QM region to one cell ensures rather limited computational costs, while the choices of QM method, population analysis, and LJ parameter set allow for a flexible model. A drawback is the description of interactions between the QM unit cell and all other cells is limited to the QM/MM level, which is not necessarily sufficient for small intermolecular distances and strong interactions. This method is further discussed in Chapter 4.1.

The Hybrid Many-body Interaction Approach

The hybrid many-body interaction (HMBI) approach²³ uses a different method to account for the higher order contributions. The truncated expansion (Equation 2.48) is applied at the QM level, but further terms are taken at the MM level. An MM description of the periodic system is used, which by construction includes all terms, so, the one- and two-body parts need to be excluded. Hence, the energy expression

$$\begin{aligned}
 E &= E_{\text{MM}} + \sum_{A \in \text{UC}} (E_{\text{QM}}(A) - E_{\text{MM}}(A)) \\
 &\quad + \frac{1}{2} \sum_{A \in \text{UC}} \sum_{B \in \text{UC}} d(r_{AB}) (\Delta E_{\text{QM}}(AB) - \Delta E_{\text{MM}}(AB)) \\
 &\quad + \sum_{A \in \text{UC}} \sum_{B \in \text{NC}} d(r_{AB}) (\Delta E_{\text{QM}}(AB) - \Delta E_{\text{MM}}(AB))
 \end{aligned} \tag{2.55}$$

uses the MM energy of the periodic system, subtracts the one- and two-body contributions at the MM level, and adds these contributions at the QM level. A damping function $d(r_{AB})$ based on the distance r_{AB} between molecules A and B ensures a smooth transition between the description of interactions at the QM and MM level. The use of polarisable force fields is required to allow for a reasonable description of the MM terms included in the HMBI energy expression. In conjunction with CCSD(T) lattice energies within chemical accuracy can be obtained.

The Hybrid QM/QM Approach

Similar to the HMBI approach, one may as well treat one- and two-body contributions at a more accurate level and higher order contributions with a cheaper QM level. Initial attempts²⁴ were based on a similar incremental scheme for the treatment of electron correlation in covalent crystals. For molecular crystals, models based on periodic HF^{53,54} or DFT^{55,56,57,58} have been reported with two-body corrections described

at different levels of theory. Very similar to Equation 2.55, the expression

$$\begin{aligned}
E &= E_{\text{Low}} + \sum_{A \in UC} (E_{\text{High}}(A) - E_{\text{Low}}(A)) \\
&\quad + \frac{1}{2} \sum_{A \in UC} \sum_{B \in UC} (\Delta E_{\text{High}}(AB) - \Delta E_{\text{Low}}(AB)) \\
&\quad + \sum_{A \in UC} \sum_{B \in NC} (\Delta E_{\text{High}}(AB) - \Delta E_{\text{Low}}(AB))
\end{aligned} \tag{2.56}$$

arises which includes the whole description at a low QM level with energy contributions E_{Low} . Again, high-level contributions E_{High} are added, while the respective low-level parts need to be subtracted from the complete low-level energy. In contrast to the HMBI scheme one can expect an improved description of the higher order terms. However, the same limitations of periodic HF or DFT calculations apply (e.g. with respect to the number of atoms per cell that can be treated).

3 Methods

3.1 Computational Details

In this chapter, a small description is given for the standard methods used in the remaining thesis. Unless otherwise noted, the basis sets and input options used throughout are detailed in the paragraphs below.

The whole model was developed within the ChemShell 3.5.0 environment,²⁵ making use of its built-in QM/MM suite.²⁶

DFT calculations were performed with the ORCA 3.0.3 programme package,⁴⁹ in conjunction with the Gennbo 5.9 programme⁵⁹ for the calculation of NPA charges.³⁸ The GGA functional PBE,⁶⁰ meta-GGA functional TPSS,⁶¹ and the hybrid functional B3LYP⁶² were used. Dispersion corrections according to the DFT-D3 scheme⁴³ with Becke-Johnson damping^{63,64,65} were used throughout all DFT calculations. The def2-TZVP and def2-QZVPP⁶⁶ basis sets were used for DFT calculations, but the def2- prefix was dropped throughout. Density fitting approximations^{67,68,69,70} to the Coulomb integrals were employed using the def2-TZVP/J, and def2-QZVPP/J⁷¹ auxiliary basis sets. Hybrid DFT calculations were performed under the COSX⁷² approximation to the exchange integrals. Chelpg charges⁷³ were calculated with the Gaussian09 programme package.⁷⁴

Wave-function calculations were performed with the MOLPRO 2012.1.23 programme package.⁷⁵ Second order Møller-Plesset perturbation theory without (MP2) and with local approximations (LMP2)^{76,77} were used as well as the spin component scaling variants,³³ denoted as SCS-MP2 and SCS-LMP2. Local coupled cluster⁷⁸ with singles and doubles,^{79,80} and non-iterative perturbative triples^{81,82,83} (LCCSD(T0)) was applied. Density fitting approximations^{67,68,69,70} for local approximations^{84,85,70} were used throughout all wave-function calculations. The cc-pVXZ ($X=T,Q$)⁸⁶ basis sets were used for all wave-function calculations. These are shortly denoted as VXZ. The respective auxiliary basis sets cc-pVXZ/JK⁸⁷ for HF and cc-pVXZ/MP2⁸⁸ for the subsequent correlated calculations were used. Local orbitals were obtained according to the Pipek-Mezey method³⁷ and a criterion of $T_{\text{NPA}} = 0.03$ for the domain selection was applied.³⁹

For ac-QM/MM calculations atomic point charges were calculated either according to the Natural Population Analysis³⁸ (NPA) or the Chelpg scheme.⁷³ Lennard-Jones (LJ) parameters for the QM/MM calculations were taken from the Universal Force Field⁸⁹ (UFF), Generalised Amber Force Field⁹⁰ (GAFF), Charmm Generalised Force Field⁹¹ (CGFF), and the MM3 force field.⁹² MM3 parameters were taken from the Tinker 7.1.2 programme⁹³ data and converted to the Exp-C6 format to be usable in ChemShell.

All geometry optimisations were performed with the dl-find optimiser.⁹⁴ Cartesian coordinates were used together with the limited-memory Broyden-Fletcher-Goldfarb-Shanno optimisation algorithm.⁹⁵ Convergence thresholds of $3 \cdot 10^{-6} E_h$ (energy), $5.4 \cdot 10^{-3}$ and $3.6 \cdot 10^{-3} a_0$ (maximum step component and RMS step), and $13.5 \cdot 10^{-4}$ and $9.0 \cdot 10^{-4} E_h \cdot a_0^{-1}$ (maximum gradient component and RMS gradient) were applied to signal convergence.

Optimiser Issues

A special issue related to the optimiser dl-find employed for all ChemShell calculations is the optimiser memory reset (OMR). In different cases and for various reasons, an erratic convergence behaviour was observed. The optimiser is hard-coded to never accept an optimisation step leading to an increase in energy, and to decrease the step size whenever the latter is encountered. After several rejections, however, the optimiser resets its memory, seemingly replacing reference coordinates and energies with new values. From

that point the optimiser internal values and the ac-QM/MM values are different and lead to an inconsistent and unpredictable behaviour in the interaction of both. If convergence had almost been achieved before the OMR, its influence might be negligible, but that can not clearly be determined. Therefore, it is not clear how much the OMR effects the results if convergence is signalled after such an event. The corresponding values are usually still presented, but marked as questionable ones.

3.2 Statistical Methods

Results obtained with the presented model are compared to experimental or theoretical reference data. Statistical values and statistical differences are calculated for sets of uniform data (e.g. a set of atomic gradients, or a set of cohesive energies for several molecules). Among a single set of N values A_i , the arithmetic mean

$$\text{mean} = \frac{1}{N} \sum_{i=1}^N A_i \quad (3.1)$$

and the root mean square value

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N A_i^2} \quad (3.2)$$

are used. When comparing two arbitrary sets of values A_i and B_i , the mean deviation

$$\text{MD} = \frac{1}{N} \sum_{i=1}^N (A_i - B_i), \quad (3.3)$$

mean absolute deviation

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N (|A_i - B_i|), \quad (3.4)$$

and root mean square deviation

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - B_i)^2} \quad (3.5)$$

are calculated.

3.3 The Benchmarking Reference

In order to have a measure for the quality and reliability of the proposed method, the X23 benchmark set was chosen. It is based on the C21 set⁹⁶ introduced by Otero-de-la-Roza and Johnson, and expanded by Reilly and Tkatchenko.⁹⁷ The X23 set consists of 23 small to medium sized molecular crystals (Figure 3.1) for which experimental structure and thermochemistry data are available. A range of crystals from van der Waals (vdW) dominated systems (e.g. benzene, #6) to strongly hydrogen bonded ones (e.g. ammonia, #4) is covered, including one case of crystal structure polymorphism (oxalic acid, #15 and #16). Experimental structures and back-corrected pure electronic cohesive energies^{96,97} are available as well as plane-wave DFT results for these values.⁹⁸

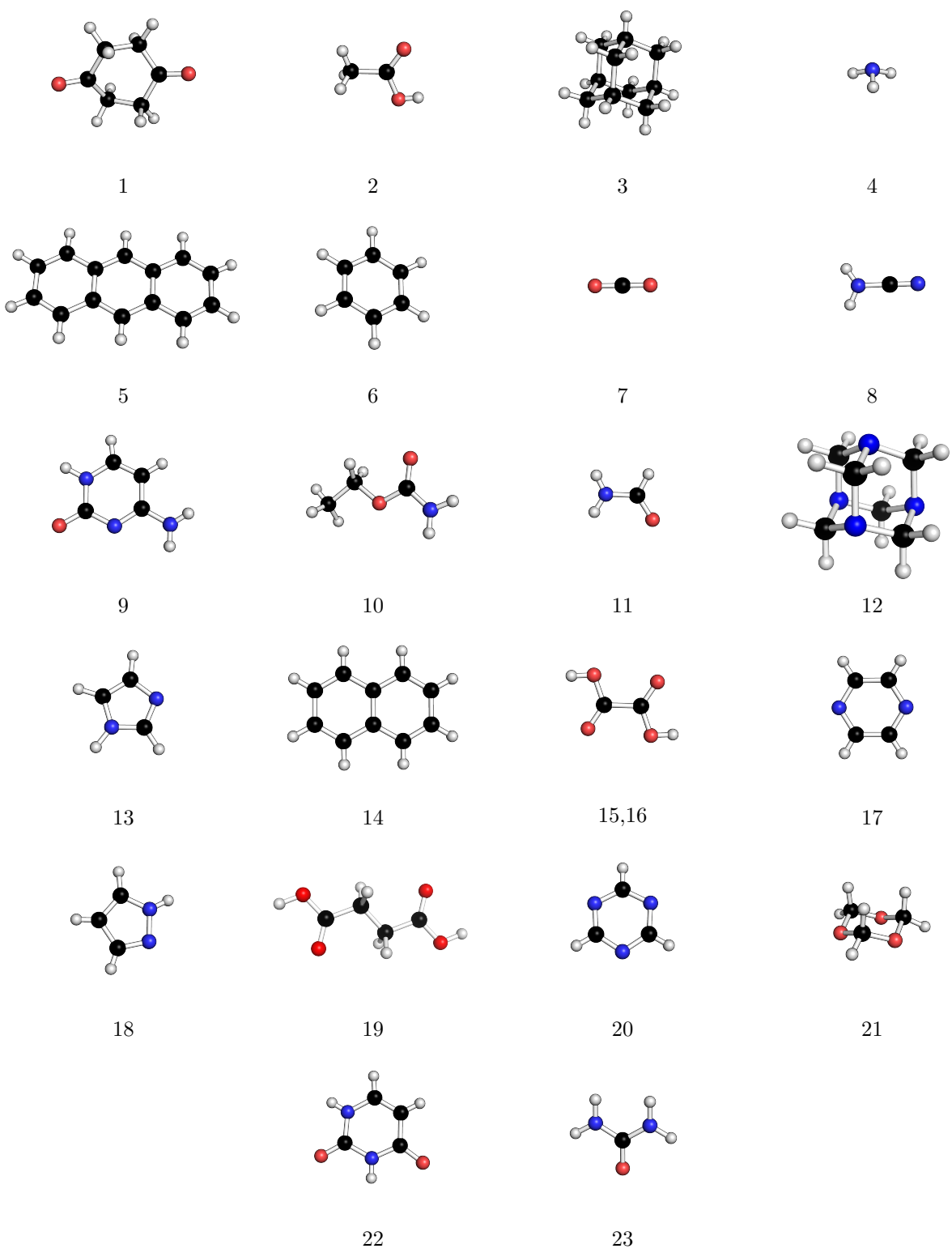


Figure 3.1: Molecules of the X23 Benchmark Set.

4 The ac-QM/MM Model

4.1 Bjornsson-Bühl QM/MM

The original scheme developed by Bjornsson and Bühl²² (BB) is used as a starting point, but some slight modifications in the coding were required to allow for the treatment of systems with more than one molecule per unit cell. All structures of the X23 benchmark set were optimised at the PBE-D3/TZVP+NPA+CGFF level of theory, using the BB scheme. The results clearly show a couple of serious flaws.

The first problem to note in the BB scheme is its energy expression. An optimisation is performed in order to minimise the QM/MM energy obtained for a QM region representing the unit cell molecules. The energy used for optimisation, however, is neither connected to the cohesive energy nor any effective crystal stabilisation.

A second problem is the definition of a QM region within the MM environment. The BB scheme uses all molecules of one cell as the QM region and for optimisation, thus, allowing for some intermolecular interactions to be calculated at the QM level. Consequently, interactions that are required to be the same due to crystal symmetry are once evaluated at the QM level, and once at the QM/MM level for two equivalent molecules. This may evoke a change to both molecular internal geometry and alignment among the molecules of the cell.

Thirdly, the BB scheme only allows for the optimisation of the molecules of one unit cell while the environment geometry is kept fixed. While for very rigid systems this may be a reasonable assumption, more flexible molecules will significantly rearrange within the static environment. Especially for common X-ray structure data, in which hydrogen atoms are usually not clearly determined, standardised (often quite short) distance parameters for bonds with hydrogen are used. The QM molecules may have a somewhat reduced contact to their environment and, therefore, even more freedom to rearrange. In that case the translational symmetry is broken and an unphysical structure is obtained. Additionally, the BB scheme does not allow for the optimisation of cell parameters.

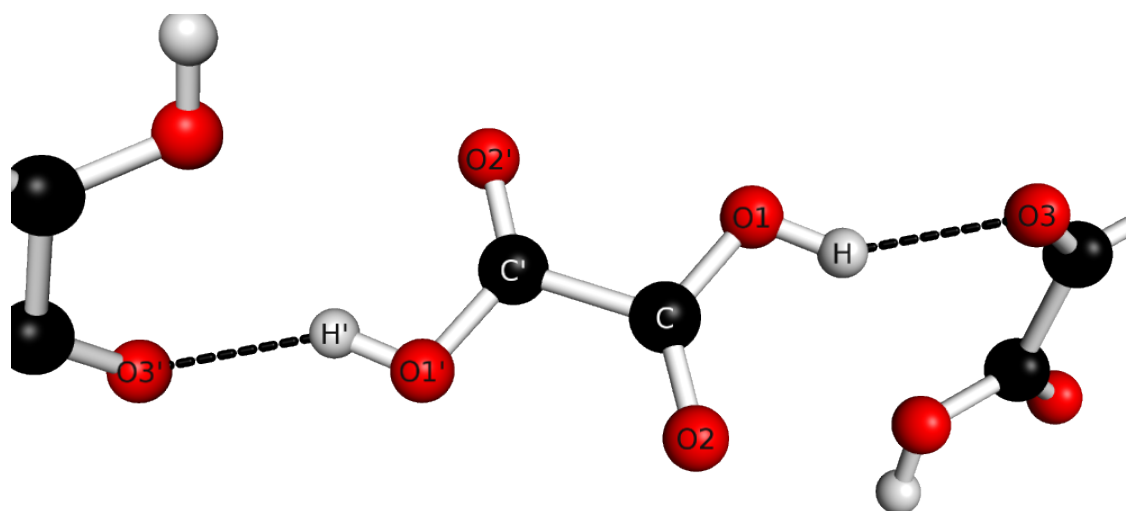


Figure 4.1: The active unit of oxalic acid (α) from an optimisation according to the BB scheme. Calculation at the PBE-D3/QZVPP+NPA+CGFF level of theory.

Table 4.1: Selected geometry parameters from a PBE-D3/QZVPP+NPA+CGFF optimisation of oxalic acid (α) according to the BB scheme. Experimental values (Exp.) and results of the four molecules in the QM region (Mol.1 to Mol. 4). Distances r in Å, dihedral angles dih in $^\circ$.

	$r(\text{O1H})$	$r(\text{O1'H1'})$	$r(\text{H}\cdots\text{O3})$	$r(\text{H}\cdots\text{O3}')$	$dih(\text{O2'C1'C1O1})$	$dih(\text{O2C1C1'O1'})$
Exp.	0.926	0.926	1.886	1.886	-0.2	0.2
Mol. 1	0.968	0.970	2.114	2.125	1.1	-0.7
Mol. 2	1.006	0.969	1.689	2.091	-10.2	-10.1
Mol. 3	0.993	0.970	1.861	2.099	2.3	1.4
Mol. 4	0.971	0.971	2.135	2.131	9.7	9.8

A prime example for both aforementioned effects is the case of oxalic acid (α) (Figure 4.1) with four molecules in the unit cell. While the intramolecular bond lengths and angles are quite similar for all four molecules, the internal dihedral angles and intermolecular geometry parameters change strongly (Table 4.1). In the case of the O-C-C-O dihedrals the results range from -10.2 to $+9.8^\circ$ (compared to values of $\pm 0.2^\circ$ from experiment), indicating a significant twist around the C-C axis. Moreover, the optimisation yields intermolecular hydrogen bonds between 1.689 and 2.135 Å for the different molecules. These results prove a severe distortion of the crystal structure between the optimised molecules and the environment, and among the optimised molecules as well.

From a technical point of view the approach is also inefficient, since both the number of atoms per molecule and the number of molecules per cell increase the size of the QM region. For a hybrid QM/MM approach to a symmetric problem (assuming the molecules of the unit cell do have a symmetry relation) it seems much more reasonable to treat one molecule at the QM level and all interactions with other molecules at the QM/MM level (or pairwise at the QM level).

For these reasons the improved ac-QM/MM scheme was developed and tested, which is presented in the following.

4.2 Overview of the ac-QM/MM Scheme

The whole model is based on the assumption that one molecule can represent the whole crystal. Hence, one QM molecule is surrounded with a cluster of MM molecules, but internal energy contributions and interactions are restricted to this single molecule. It is, therefore, referred to as the central molecule throughout. All other molecules are denoted as cluster molecules, environment molecules, or neighbour molecules. The overall scheme represents an additive crystal QM/MM model and is referred to as ac-QM/MM. The general procedure is presented in Figure 4.2. Its parts are further explained in the following sections.

Crystal structure information is used to construct the finite cluster around the central molecule. Cluster molecules are selected according to their distance from the cluster origin. Symmetry relations between these molecules are calculated. These serve to identify equivalent atoms which must behave accordingly. However, intramolecular symmetry is neglected. The procedures are described in detail in Section 4.4.1. LJ parameters from a force field are assigned to all cluster atoms, these are not changed during the optimisation. A set of initial charges is calculated for the central molecule and assigned to all cluster atoms according to the symmetry relations. Whenever the theory level of an ac-QM/MM calculation is stated, the population analysis and force field are appended to the QM theory definition. A calculation PBE-D3/TZVP+NPA+CGFF would, therefore, refer to an ac-QM/MM calculation at the PBE-D3/TZVP level of theory for the QM part with NPA charges and CGFF parameters used for the QM/MM interactions.

For each structure the initial step is a self-consistent calculation of charges, referred to as the charge self-consistent field (QSCF) cycle. QM charges $q(\text{QM})$ are determined in the electrostatic potential of the cluster. These charges are compared to the current charges $q(i)$. If the difference does not fulfil a convergence criterion, the charges $q(\text{QM})$ are assigned to the central molecule and the cluster molecules, and saved as $q(i)$. Background charges are fitted and placed at the cluster surface. If the convergence criterion is fulfilled, the charges $q(\text{QM})$ are discarded and the charges $q(i)$ are possibly scaled or damped. The QSCF procedure is described in detail in Section 4.6.

For the set of converged charges, gradients are calculated. Gradients are composed from atomic gradients of both the central molecule and the environment molecules. An optimisation step for the central molecule is formed by an external optimiser and applied. All cluster molecules and cell parameters are updated accordingly. The gradient calculation and geometry update procedure is described in detail in Section 4.4. If the optimisation step is accepted, the current geometry $r(i)$ and charges $q(i)$ are saved as the accepted geometry $r(\text{acc})$ and charges $q(\text{acc})$ for the next step, otherwise the changes are discarded.

Energy and gradients are always initially calculated at the one-body level (1B), but two-body corrections (2B) can be added. In the former case, only the central molecule is calculated at the QM level, all intermolecular interactions are calculated at the QM/MM level. At the two-body level, pairwise interactions between molecules are calculated at the QM level. The respective formulae are presented in detail in Section 4.7.

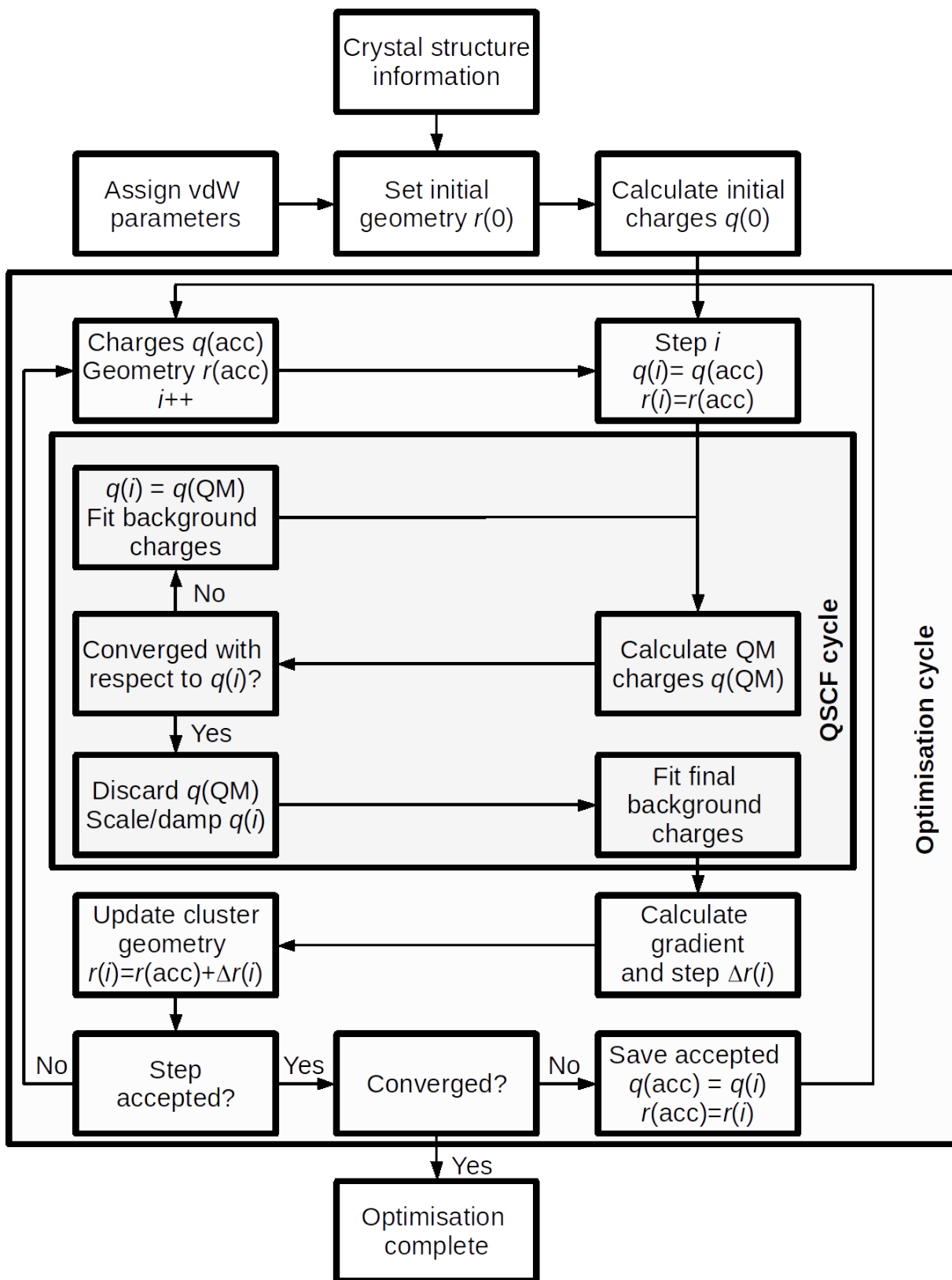


Figure 4.2: Schematic representation of an ac-QM/MM geometry optimisation.

4.3 A Consistent Energy Expression

A crucial point is the definition of a consistent energy for a full system optimisation. As a single molecule is used to represent the whole crystal, such an energy expression must represent the crystal energy E_{cry} adequately. The whole crystal may be considered as a system of M single molecules m that all have an internal energy $E(m)$ and interactions $\Delta E(mn)$ with all other molecules n . The crystal energy per molecule is then

$$E_{\text{cry}} = \frac{1}{M} \left(\sum_m^M E(m) + \frac{1}{2} \sum_m^M \sum_{n \neq m}^M \Delta E(mn) \right). \quad (4.1)$$

Since we require all molecules to be the same, the expression reduces to

$$E_{\text{cry}} = E(0) + \frac{1}{2} \sum_{n \neq 0}^M \Delta E(0n). \quad (4.2)$$

Our single molecule calculation of the central molecule 0, on the other hand, provides an embedded energy

$$E_{\text{emb}} = E(0) + \sum_m^M \Delta E(0m). \quad (4.3)$$

It consists of the internal energy $E(0)$ and the interaction energies $\Delta E(0m)$ with all other molecules m . To avoid double-counting of the intermolecular interactions, the crystal energy is calculated as

$$E_{\text{cry}} = \frac{1}{2} (E_{\text{emb}}(0) + E(0)). \quad (4.4)$$

The embedded energy $E_{\text{emb}}(0)$ results from the fully embedded QM/MM calculation, the internal energy $E(0)$ is taken from a gas phase calculation of molecule 0 at its embedded geometry. Then the stabilisation of a single molecule in the crystal environment compared to the gas phase is the cohesive energy

$$E_{\text{coh}} = E_{\text{cry}} - E_{\text{gp}}, \quad (4.5)$$

where E_{gp} is the energy of a single molecule optimised in the gas phase. While the underlying embedded energy can be calculated according to different schemes, cohesive energies are always computed according to Equation (4.5).

At the one-body level the embedded energy is simply the sum of the QM energy of the central molecule in the field of the surrounding point charges and the interaction energy calculated between the central molecule and all other cluster molecules through the LJ pair potentials. It is denoted as ac(1)-QM/MM in the following and simply provided by the energy term

$$E_{\text{ac}(1)\text{-QM/MM}} \equiv E_{\text{emb}}^{\text{1B}}. \quad (4.6)$$

At the two-body level, more complicated energy expressions are required. These are presented in the respective Section 4.7.

A final note on the cohesive energies should be given. Comparison is made to experimental values which are back-corrected sublimation enthalpies. It is, unfortunately, not necessarily clear how the gas phase state of a molecule looks like. If the gas phase optimisation does not provide the correct minimum, the cohesive energy will be shifted accordingly. Moreover, the assumption of E_{coh} uses a single optimised molecule. However, experimental energies are derived from sublimation enthalpies, but these might as well measure sublimation to gas phase dimers. If the experimental sublimation yields gas phase dimers, the back-corrected cohesive energy will differ from the theoretical value

4.4 Generalised Geometry Gradients and Crystal Optimisation

4.4.1 ac-QM/MM Internal Gradients

General Approach

One step to eliminate the deficiencies of the original BB scheme is the introduction of an environment update. This requires the consideration of crystal symmetry to ensure the geometry of all molecules is changed accordingly. While translational symmetry can be included easily, the representation of rotations, reflections and inversions is more cumbersome.

To account for symmetry between molecules, the following limited symmetry approach was made. The internal symmetry of molecules is neglected, which means that e.g. for a carbon dioxide molecule formally both oxygen atoms are not the same. However, it is expected that due to the symmetric environment equivalent atoms will have almost the same interactions and, therefore, behave in the same way. Intermolecular symmetry, on the other hand, may not simply be neglected, as the QM unit shall be restricted to a single molecule. It is handled in a simplified manner, extracting minimal symmetry information from atomic coordinates and connectivity between atoms. Hence, it is not limited to explicit symmetry definitions, e.g. from a crystallographic information file.

Each cluster molecule m consists of atoms A with cartesian coordinates $\vec{r}(m, A)$. Due to the crystal symmetry, there are symmetry operations which can generate each cluster molecule m from the central molecule 0 as

$$\vec{r}(m, A) = \underline{U}(m)\vec{r}(0, A) + \vec{t}(m). \quad (4.7)$$

All atom coordinates of a molecule m are generated according to a combination of rotation \underline{U} and translation \vec{t} . Since symmetry is used to transform coordinate changes and atom gradients, the translation component is not required and does not show up in the respective equations. It remains

$$\Delta\vec{r}(m, A) = \underline{U}(m)\Delta\vec{r}(0, A) \quad (4.8)$$

for the transformation of a coordinate displacement $\Delta\vec{r}$. While for highly symmetric space groups there can be a large amount of symmetry operations to generate the same molecule m from the central molecule 0, we basically need only one of them to apply coordinate changes from one molecule to the other. Therefore, a rather simple approach, which does not require an explicit definition of symmetry, is chosen to determine the matrix \underline{U} from atomic coordinates.

Determination of the Rotation Matrix

The approach to determine \underline{U} is based on equation 4.7. Since the same transformation is applied to all atoms, it must be possible to determine \underline{U} from a limited sample. Four non-coplanar atoms can be used to define a coordinate system for molecule m . After rotation according to equation 4.7, the same construction yields a rotated coordinate system. Step-by-step alignment of the axes is used to determine \underline{U} . Basically, the three axes $\vec{a}_i(m)$ ($i \in \{1, 2, 3\}$) of molecule m are determined as

$$\vec{a}_i^{(0)}(m) = \vec{r}_{i+1}(m) - \vec{r}_1(m) \quad (4.9)$$

from the coordinates \vec{r} of the first to fourth atoms. The superscript indicates the step of the alignment procedure, i.e. the (0) indicates the initial vectors. If these vectors are coplanar or collinear, further atoms are used instead of atoms 3 and 4, until a set of non-coplanar axes is obtained. Otherwise the molecule is determined as planar or linear. In these cases only two (planar) or one (linear) operations are required to align the axes.

The alignment procedure for a non-coplanar set of axes then uses three subsequent steps, each providing

a single transformation matrix $\underline{\underline{U}}^{(i)}$ ($i \in \{1, 2, 3\}$). This allows for the calculation of the final $\underline{\underline{U}}$ matrix as

$$\underline{\underline{U}} = \underline{\underline{U}}^{(3)}\underline{\underline{U}}^{(2)}\underline{\underline{U}}^{(1)}. \quad (4.10)$$

If $\vec{a}_1^{(0)}(0)$ and $\vec{a}_1^{(0)}(m)$ are already parallel, $\underline{\underline{U}}^{(1)}$ is a unit matrix, otherwise it is calculated as a rotation matrix. An auxiliary axis

$$\vec{\sigma} = \frac{\vec{a}_1^{(0)}(0) \times \vec{a}_1^{(0)}(m)}{|\vec{a}_1^{(0)}(0) \times \vec{a}_1^{(0)}(m)|} \quad (4.11)$$

is constructed and used as a rotation axis between $\vec{a}_1^{(0)}(m)$ and $\vec{a}_1^{(0)}(0)$. The matrix $\underline{\underline{R}}$ associated to an arbitrary rotation by an angle α around a normalised vector \vec{n} is

$$\underline{\underline{R}} = \begin{pmatrix} n_x^2(1 - \cos(\alpha)) + \cos(\alpha) & n_x n_y(1 - \cos(\alpha)) - n_z \sin(\alpha) & n_x n_z(1 - \cos(\alpha)) + n_y \sin(\alpha) \\ n_y n_x(1 - \cos(\alpha)) + n_z \sin(\alpha) & n_y^2(1 - \cos(\alpha)) + \cos(\alpha) & n_y n_z(1 - \cos(\alpha)) - n_x \sin(\alpha) \\ n_z n_x(1 - \cos(\alpha)) - n_y \sin(\alpha) & n_z n_y(1 - \cos(\alpha)) + n_x \sin(\alpha) & n_z^2(1 - \cos(\alpha)) + \cos(\alpha) \end{pmatrix}. \quad (4.12)$$

The first rotation matrix $\underline{\underline{U}}^{(1)}$ is calculated according to equation 4.12 from the angle between $\vec{a}_1^{(0)}(0)$ and $\vec{a}_1^{(0)}(m)$. It is applied to all axes of molecule 0,

$$\vec{a}_i^{(1)}(0) = \underline{\underline{U}}^{(1)}\vec{a}_i^{(0)}(0) \quad \forall i \in \{1, 2, 3\} \quad (4.13)$$

The axes of m are not changed,

$$\vec{a}_i^{(1)}(m) = \vec{a}_i^{(0)}(m) \quad \forall i \in \{1, 2, 3\}. \quad (4.14)$$

In a second step, $\vec{a}_2^{(1)}(0)$ is aligned to $\vec{a}_2^{(1)}(m)$. The $\vec{a}_1^{(1)}(0)$ component is projected out of $\vec{a}_2^{(1)}(0)$ and $\vec{a}_2^{(0)}(m)$ as

$$\vec{a}_2^{(1)}(i) = \vec{a}_2^{(1)}(i) - \vec{a}_1^{(1)}(i) [\vec{a}_1^{(1)}(i) \cdot \vec{a}_2^{(1)}(i)] \quad \forall i \in \{0, m\}. \quad (4.15)$$

Now $\vec{a}_1^{(1)}(0)$ must be orthogonal to $\vec{a}_2^{(1)}(0)$ and $\vec{a}_2^{(1)}(m)$. Therefore, it can be used for the rotation between the latter axes. The second rotation matrix $\underline{\underline{U}}^{(2)}$ is calculated from $\vec{a}_1^{(1)}(0)$ and the angle between $\vec{a}_2^{(1)}(0)$ and $\vec{a}_2^{(1)}(m)$ according to Equation 4.12. The application to the axes $\vec{a}^{(1)}$ yields

$$\vec{a}_i^{(2)}(0) = \underline{\underline{U}}^{(2)}\vec{a}_i^{(1)}(0) \quad \forall i \in \{1, 2, 3\}. \quad (4.16)$$

Again, the axes of m are not changed,

$$\vec{a}_i^{(2)}(m) = \vec{a}_i^{(1)}(m) \quad \forall i \in \{1, 2, 3\}. \quad (4.17)$$

If both axes have been aligned successfully, the third axes may either be the same, then $\underline{\underline{U}}^{(3)}$ is a unit matrix, or a reflection at the plane spanned by $\vec{a}_1^{(2)}(0)$ and $\vec{a}_2^{(2)}(0)$ is required. In the latter case the reflection is performed in a three step procedure, yielding the transformation matrix $\underline{\underline{U}}^{(3)}$ as

$$\underline{\underline{U}}^{(3)} = \underline{\underline{A}}^{-1}\underline{\underline{R}}\underline{\underline{A}}. \quad (4.18)$$

$\underline{\underline{A}}$ is an auxiliary alignment matrix calculated for a rotation of axis $\vec{a}_3^{(2)}(0)$ onto the z-axis. An auxiliary axis and the angle between $\vec{a}_3^{(2)}(0)$ and the z-axis are used to calculate $\underline{\underline{A}}$ according to 4.12. Subsequently, a reflection is performed on the xy-plane using

$$\underline{\underline{R}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (4.19)$$

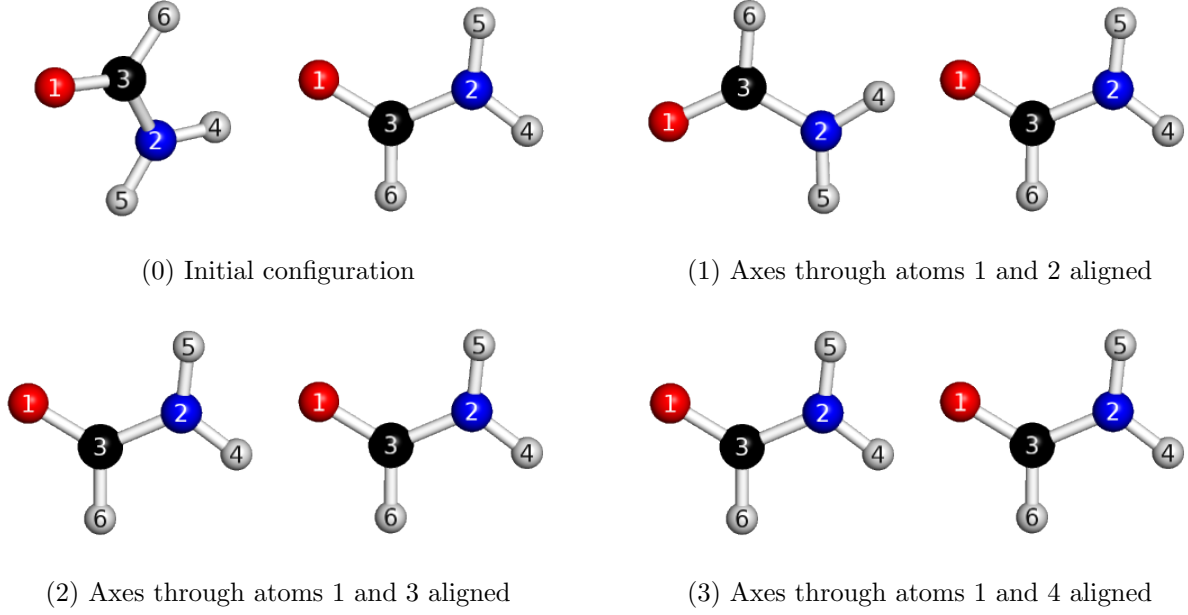


Figure 4.3: Step-by-step alignment of the left formamide molecule to the right one following the ac-QM/MM procedure. Three steps from the initial configuration to the fully aligned final configuration are performed. The numbering from (0) to (3) refers to the definition of molecular axes during the alignment procedure (Equations 4.9 to 4.20).

Finally, a rotation with the inverse of \underline{A} is performed to restore the previous orientation. After the application of $\underline{U}^{(3)}$,

$$\underline{a}_i^{(3)}(0) = \underline{U}^{(3)} \underline{a}_i^{(2)}(0) \quad \forall i \in \{1, 2, 3\}, \quad (4.20)$$

all axes $a(0)$ and $a(m)$ are aligned.

Although the occurring numbers are checked repeatedly to avoid problems with numerical accuracy, e.g. for dot products being not exactly 0 due to the limited accuracy of coordinate values, it cannot be guaranteed that the scheme does always work as desired. Possibly changes to the thresholds or additional checks of numbers might be required, but no problems could be observed for all systems investigated in this work.

Since the geometry update is applied to all cluster atoms, the underlying gradients have to be modified accordingly. Hence, gradient contributions for the displacement of all non-QM cluster atoms need to be included to form reasonable atomic ac-QM/MM gradients. These consist of LJ gradients and point charge gradients, which are both readily available from the QM/MM calculation. These are combined according to an expression of the form

$$\underline{g}(A) = \left(\nabla(0, A) - \frac{1}{N} \sum_{B \in 0} \nabla(0, B) + \sum_{m \neq 0} \underline{U}^{-1}(m) \left(\nabla(m, A) - \frac{1}{N} \sum_{B \in m} \nabla(m, B) \right) \right) E_{\text{cry}}. \quad (4.21)$$

The \underline{U}^{-1} are the inverse matrices of the rotation matrices \underline{U} and the operator $\nabla(m, X)$ is the cartesian gradient operator for atom X in molecule m . Since translational movement of any molecule m is not intended to arise from these gradients, a respective contribution needs to be removed from the atomic gradients. This leads to operators of the form

$$\nabla_{\text{eff}}(m, A) = \nabla(m, A) - \frac{1}{N} \sum_{B \in m} \nabla(m, B) \quad (4.22)$$

in Equation 4.21. The sum over all cluster molecules m in Equation 4.21 accounts for the contribution of all equivalent atoms, aligned by the rotation matrices $\underline{U}^{-1}(m)$.

Relevance of the MM Gradient Contributions

The update procedure according to Equation 4.8 was implemented and ac(1)-QM/MM geometry optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory were performed for several of the X23 structures together with simple QM/MM optimisations of only one molecule in a frozen environment. The single environment contributions are notably smaller than the QM/MM gradients of the central molecule, but a large number of the former is added. Hence, it is not clear how much the contributions add up or cancel out, in which case they might be negligible. Therefore, for several systems calculations were performed in the following ways:

- QM/MM: Optimisation of the central molecule with the default QM/MM gradients
- QM/MM+EnvUp: Optimisation of the central molecule with the default QM/MM gradients, update of all other molecular geometries accordingly
- ac(1)-QM/MM: Optimisation of the central molecule with ac(1)-QM/MM gradients, update of all other molecular geometries accordingly

The convergence behaviour of these calculations in Table 4.2 shows the need for the use of ac(1)-QM/MM gradients in conjunction with the environment update. Both simple QM/MM and ac(1)-QM/MM optimisations converge very smoothly with an average step acceptance rate of over 90%. Only very few calculations display a really slow convergence, e.g. for pyrazine the QM/MM optimisation required 110 steps until convergence was reached. Nevertheless, reasonable mean numbers of 28.0 and 21.9 steps were obtained for the simple QM/MM and ac(1)-QM/MM approaches, respectively. Calculations using an environment update in conjunction with the unchanged QM/MM gradients, however, do not converge at all. For the only case of carbon dioxide a converged result was obtained, the ten residual structures all suffered from an OMR (Section 3.1) after several rejected optimisation steps.

An individual consideration of the respective atomic contributions from both the central molecule and its closest six neighbours provides some insight to the underlying reasons. From the QM/MM gradients in Table 4.3 one can see that the gradients from the MM region (point charge gradients plus LJ contribution from the QM/MM interaction with the QM region) are significantly smaller than those from the QM region (QM gradients plus LJ contribution from the QM/MM interaction with the environment). The latter are of the order of $3 \cdot 10^{-2} E_h \cdot a_0^{-1}$, while the former are all smaller than $0.86 \cdot 10^{-2} E_h \cdot a_0^{-1}$. For more distant molecules even smaller gradients are obtained, hence, the combined ac(1)-QM/MM gradients are still not too different from the unmodified QM/MM ones.

This seems to be a negligible maximum difference, which does not explain the convergence problems with unmodified QM/MM gradients, but the situation clearly changes during the optimisation. After nine optimisation steps the gradient components of the QM region have decreased much more than those of the MM region, now values in the order of $10^{-4} E_h \cdot a_0^{-1}$ (see Table 4.4) are obtained.

Consequently, the impact of the MM contributions to the ac(1)-QM/MM gradient significantly increases, leading to large deviations between the QM/MM and ac(1)-QM/MM gradients of the central molecule. In fact, 6 of 12 ac(1)-QM/MM gradient components (which were calculated, but not applied) do not even show the same algebraic sign as the QM/MM ones, and differences of up to $0.62 \cdot 10^{-2} E_h \cdot a_0^{-1}$ are found. The absolute differences are comparably large to those of the starting structure, but due to the smaller gradient values the percentage differences of ac(1)-QM/MM with respect to unmodified QM/MM reach more than 1000% (e.g. for g_y on H(4)). Obviously the environment contributions are only negligible in the beginning of the optimisation, but gain great importance once the gradients become smaller. Hence, the environment update based on simple QM/MM gradients fails at last once an optimisation approaches convergence and cause the observed erratic convergence behaviour.

Interestingly, the QM/MM gradients seem to be larger after the same number of optimisation steps with ac(1)-QM/MM. As can be seen from Table 4.5, several gradient contributions from both the QM and MM

Table 4.2: Convergence behaviour of ac(1)-QM/MM optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory. Either unmodified QM/MM gradients are used to optimise only the central molecule (QM/MM), or unmodified QM/MM gradients are used to optimise the central molecule and update all other molecules accordingly (QM/MM+EnvUp), or the ac(1)-QM/MM procedure is used (ac(1)-QM/MM). The numbers of accepted steps N_{acc} and total steps N_{tot} until convergence is achieved or the step limit of 200 is reached are given. Due to an OMR during some calculations convergence was signalled but not necessarily achieved, or not reached within the limit of 200 steps. The respective results are enclosed in round brackets.

	QM/MM		QM/MM+EnvUp		ac(1)-QM/MM	
	N_{acc}	N_{tot}	N_{acc}	N_{tot}	N_{acc}	N_{tot}
1,4-cyclohexanedione	31	32	–	–	30	31
acetic acid	31	32	(21)	(200)	19	20
adamantane	18	18	–	–	16	17
ammonia	14	15	(19)	(200)	7	7
anthracene	31	32	–	–	29	32
benzene	8	8	–	–	7	8
carbon dioxide	4	4	5	5	4	4
cyanamide	25	27	(16)	(200)	19	20
cytosine	38	40	–	–	45	45
ethylcarbamate	41	41	–	–	43	48
formamide	18	19	(35)	(200)	30	30
hexamine	8	9	–	–	8	20
imidazole	24	25	(8)	(23)	11	12
naphthalene	35	37	–	–	30	33
oxalic acid (α)	22	24	(17)	(200)	24	24
oxalic acid (β)	18	18	(17)	(200)	21	22
pyrazine	104	110	–	–	12	12
pyrazole	35	36	(14)	(29)	29	30
succinic acid	21	25	–	–	15	15
triazine	8	8	(8)	(21)	8	9
trioxane	25	25	–	–	19	20
uracil	35	38	–	–	31	32
urea	20	22	(18)	(200)	13	13
Mean	26.7	28.0	16.2	134.4	20.4	21.9

region exceed the convergence criterion of $0.00135 \text{ E}_h \cdot a_0^{-1}$, but the combined ac(1)-QM/MM gradients do not. Obviously the contributions from both regions finally counterbalance each other. This may be seen as a correction to the behaviour in a QM/MM scheme discarding the environment update. In the latter case, one molecule tries to adopt a favourable position in its frozen environment. This represents an artificial stabilisation from the internal energy and interactions with the frozen surrounding potentials. It is counterbalanced by a destabilisation resulting from the update of these potentials. This indicates that the environment update does improve the description, leading to a better representation of the experimental structure.

Validation of Numerical ac(1)-QM/MM Gradients

In order to assess the reliability of the numerical gradients, the coordinate displacement Δr for the gradient calculation was varied between 0.1 and $0.001 a_0$. Calculations were performed for urea, the structure showing the largest deviations, and benzene as a non-polar counterpart. Differences between the numerical and analytical gradients were calculated for the different Δr values. The results (Figure 4.4) indicate that the choice of Δr does have an impact on the numerical gradients, but the differences are rather small, especially when decreasing the Δr parameter. While it is not clear whether results for $\Delta r = 0.001 a_0$ are more precise than for the default of $\Delta r = 0.01 a_0$, the RMSDs to analytical gradients are essentially the same for both systems. At the initial as well as at the final structure the respective RMSDs change by less than 2% when Δr is varied between 0.01 and $0.001 a_0$. For displacements of more than $0.02 a_0$, the RMSDs increase by up to one order of magnitude. This is most likely a region of too large displacements which result in inaccurate numerical gradients, thus, a $\Delta r \leq 0.02 a_0$ seems advisable to obtain consistent numerical gradients. Nonetheless, for these displacement values RMSDs up to $2.8 \cdot 10^{-4} \text{ E}_h \cdot a_0^{-1}$ are observed between numerical and analytical gradients. As these are significantly smaller than the target value of $9 \cdot 10^{-4} \text{ E}_h \cdot a_0^{-1}$

Table 4.3: QM/MM gradients of the central molecule (atoms N(1) to H(4)) and the six nearest neighbour molecules in the ammonia crystal, and the resulting ac(1)-QM/MM gradients. Results from an ac(1)-QM/MM calculation at the PBE-D3/TZVP+NPA+CGFF level of theory at the experimental structure. All values in $E_h \cdot a_0^{-1}$.

Atom	QM/MM				ac(1)-QM/MM			
	g_x	g_y	g_z	$ \vec{g} $	g_x	g_y	g_z	$ \vec{g} $
N(1)	0.0194	0.0194	0.0194	0.0337	0.0232	0.0232	0.0232	0.0402
H(2)	-0.0252	-0.0081	0.0157	0.0307	-0.0276	-0.0103	0.0147	0.0329
H(3)	-0.0081	0.0157	-0.0252	0.0307	-0.0103	0.0147	-0.0276	0.0329
H(4)	0.0157	-0.0252	-0.0081	0.0307	0.0147	-0.0276	-0.0103	0.0329
N(5)	-0.0018	-0.0047	0.0012	0.0052	-	-	-	-
H(6)	0.0002	0.0008	-0.0001	0.0009	-	-	-	-
H(7)	0.0014	0.0014	0.0008	0.0021	-	-	-	-
H(8)	0.0005	0.0009	-0.0009	0.0013	-	-	-	-
N(9)	-0.0047	0.0012	-0.0018	0.0052	-	-	-	-
H(10)	0.0014	0.0008	0.0014	0.0021	-	-	-	-
H(11)	0.0009	-0.0009	0.0005	0.0013	-	-	-	-
H(12)	0.0008	-0.0001	0.0002	0.0009	-	-	-	-
N(13)	0.0041	0.0058	-0.0049	0.0086	-	-	-	-
H(14)	-0.0010	-0.0015	0.0004	0.0019	-	-	-	-
H(15)	-0.0040	-0.0046	0.0039	0.0072	-	-	-	-
H(16)	-0.0009	-0.0006	0.0012	0.0016	-	-	-	-
N(17)	0.0058	-0.0049	0.0041	0.0086	-	-	-	-
H(18)	-0.0046	0.0039	-0.0040	0.0072	-	-	-	-
H(19)	-0.0006	0.0012	-0.0009	0.0016	-	-	-	-
H(20)	-0.0015	0.0004	-0.0010	0.0019	-	-	-	-
N(21)	0.0012	-0.0018	-0.0047	0.0052	-	-	-	-
H(22)	-0.0009	0.0005	0.0009	0.0013	-	-	-	-
H(23)	-0.0001	0.0002	0.0008	0.0009	-	-	-	-
H(24)	0.0008	0.0014	0.0014	0.0021	-	-	-	-
N(25)	-0.0049	0.0041	0.0058	0.0086	-	-	-	-
H(26)	0.0012	-0.0009	-0.0006	0.0016	-	-	-	-
H(27)	0.0004	-0.0010	-0.0015	0.0019	-	-	-	-
H(28)	0.0039	-0.0040	-0.0046	0.0072	-	-	-	-

for the RMS of the atomic gradients, they are probably negligible. Nonetheless, some further investigation on the source of these differences was carried out.

In order to detect a possible underlying inconsistency in the combination of the analytical gradients, their impact on the final gradients was checked. For urea and benzene the analytical and numerical gradients were calculated both at the initial and final structure with four combinations of QM and MM contributions

- QM: Calculation without point charges and without LJ potentials
- QM+PC: Calculation with point charges, but without LJ potentials
- QM+LJ: Calculation without point charges, but with LJ potentials
- Full: Calculation with both point charges and LJ potentials

The resulting RMSDs are shown in Table 4.6 and clearly indicate the main source of the differences between numerical and analytical gradients. Even without any gradient contribution from the environment, the RMSD is already close to the value of the combined ac-QM/MM gradient. For both benzene and urea, and at both the initial and final geometry, a change in the RMSD of less than 7% is observed when environment point charges are added. The addition of LJ parameters does not invoke a notable change in the difference between analytical and numerical gradient. Therefore, it is safe to assume that the differences mainly stem from the underlying QM gradients and are not caused by the construction of the ac-QM/MM gradients.

Validation of Analytical ac(1)-QM/MM Gradients

It has been shown in Section 4.4.1 that the consideration of contributions from the environment update is crucial in order to generate reasonable gradients. The combination procedure seems to be working well

Table 4.4: QM/MM gradients of the central molecule (atoms N(1) to H(4)) and the six nearest neighbour molecules in the ammonia crystal, and the resulting ac(1)-QM/MM gradients. Results are taken from the last accepted structure of a QM/MM+EnvUp geometry optimisation at the PBE-D3/TZVP+NPA+CGFF level of theory. All values in $E_h \cdot a_0^{-1}$.

Atom	QM/MM				ac(1)-QM/MM			
	g_x	g_y	g_z	$ \vec{g} $	g_x	g_y	g_z	$ \vec{g} $
N(1)	-0.0003	0.0015	-0.0040	0.0043	0.0059	0.0077	0.0022	0.0099
H(2)	-0.0005	-0.0002	0.0003	0.0006	-0.0043	-0.0027	0.0005	0.0051
H(3)	0.0011	-0.0016	0.0033	0.0038	-0.0014	-0.0015	-0.0006	0.0021
H(4)	-0.0003	0.0003	0.0004	0.0006	-0.0001	-0.0035	-0.0021	0.0041
N(5)	-0.0009	-0.0025	0.0005	0.0027	-	-	-	-
H(6)	0.0001	0.0007	-0.0002	0.0007	-	-	-	-
H(7)	0.0007	0.0010	0.0003	0.0013	-	-	-	-
H(8)	0.0002	0.0007	-0.0006	0.0010	-	-	-	-
N(9)	-0.0024	0.0005	-0.0010	0.0027	-	-	-	-
H(10)	0.0010	0.0003	0.0007	0.0013	-	-	-	-
H(11)	0.0007	-0.0006	0.0002	0.0010	-	-	-	-
H(12)	0.0007	-0.0002	0.0001	0.0007	-	-	-	-
N(13)	0.0024	0.0032	-0.0028	0.0049	-	-	-	-
H(14)	-0.0003	-0.0007	0.0001	0.0008	-	-	-	-
H(15)	-0.0019	-0.0023	0.0021	0.0036	-	-	-	-
H(16)	-0.0002	-0.0002	0.0005	0.0006	-	-	-	-
N(17)	0.0032	-0.0028	0.0024	0.0049	-	-	-	-
H(18)	-0.0023	0.0021	-0.0019	0.0036	-	-	-	-
H(19)	-0.0002	0.0005	-0.0002	0.0006	-	-	-	-
H(20)	-0.0008	0.0001	-0.0003	0.0008	-	-	-	-
N(21)	0.0005	-0.0009	-0.0025	0.0027	-	-	-	-
H(22)	-0.0006	0.0002	0.0007	0.0010	-	-	-	-
H(23)	-0.0002	0.0001	0.0007	0.0007	-	-	-	-
H(24)	0.0003	0.0007	0.0010	0.0013	-	-	-	-
N(25)	-0.0028	0.0024	0.0032	0.0049	-	-	-	-
H(26)	0.0005	-0.0002	-0.0002	0.0006	-	-	-	-
H(27)	0.0001	-0.0003	-0.0007	0.0008	-	-	-	-
H(28)	0.0021	-0.0019	-0.0023	0.0036	-	-	-	-

for the previous tests. Nevertheless, it is deemed necessary to further assess the quality and reliability of the combined ac(1)-QM/MM gradients. Stable numerical gradients had already been obtained (Section 4.4.1), therefore, it seemed appropriate to validate the analytically calculated gradients with their numerical counterparts. Hence, numerical and analytical gradients were calculated for all X23 molecules at the PBE-D3/TZVP+NPA+CGFF level of theory. Calculations were performed on the starting structure as well as the final structure of an PBE-D3/TZVP+NPA+CGFF geometry optimisation.

At the initial structure the analytical composition of gradient contributions at the ac(1)-QM/MM level provides reasonable results. The overall RMSD between numerical and analytical gradients is in the order of $10^{-4} E_h \cdot a_0^{-1}$ for all systems, with a maximum deviation of $1.6 \cdot 10^{-4} E_h \cdot a_0^{-1}$ for urea and a minimum value of $1.6 \cdot 10^{-8} E_h \cdot a_0^{-1}$ for carbon dioxide. Except for carbon dioxide all values lie in the narrow interval between $4.1 \cdot 10^{-5}$ and $1.6 \cdot 10^{-4}$. The RMSD over the whole X23 set is $1.0 \cdot 10^{-4} E_h \cdot a_0^{-1}$. This very uniform behaviour for the whole test set indicates a general reliability of the analytical gradients used for the ac-QM/MM method. However, more notable differences are observed for the gradients of the optimised structure. Comparison of these values (Figure 4.5) shows a very consistent behaviour of the analytical gradients, providing almost the same overall RMSD ($1.0 \cdot 10^{-4} E_h \cdot a_0^{-1}$) as before. The largest and smallest deviations ($2.1 \cdot 10^{-4}$ and $3.4 \cdot 10^{-6} E_h \cdot a_0^{-1}$) are found for urea and carbon dioxide, thus, all deviations are in essentially the same range as for the initial structures. Nevertheless, the RMS gradient values are naturally much smaller for the converged than for the initial structures. In fact, these span a range between $4.2 \cdot 10^{-4}$ and $9.9 \cdot 10^{-6} E_h \cdot a_0^{-1}$ (uracil and urea) which is not much smaller than the RMSD between numerical and analytical gradients. Exemplary atom-wise comparison of the gradients indicates that most components agree quite well. Only few values differ rather strongly and thereby dominate the overall RMSD.

Nevertheless, one still lacks a contribution to the ac-QM/MM gradients. The charges are updated after each optimisation step, which includes the changes of their coordinates and magnitude. Contributions of the

Table 4.5: QM/MM gradients of the central molecule (atoms N(1) to H(4)) and the six nearest neighbour molecules in the ammonia crystal, and the resulting ac(1)-QM/MM gradients constructed from the QM/MM gradients. Results from the structure optimised at the ac(1)-QM/MM PBE-D3/TZVP+NPA+CGFF level of theory. All values in $E_h \cdot a_0^{-1}$.

Atom	QM/MM				ac(1)-QM/MM			
	g_x	g_y	g_z	$ \vec{g} $	g_x	g_y	g_z	$ \vec{g} $
N(1)	-0.0060	-0.0060	-0.0060	0.0104	0.0000	0.0000	0.0000	0.0000
H(2)	0.0036	0.0034	0.0010	0.0051	-0.0002	0.0003	-0.0002	0.0004
H(3)	0.0034	0.0010	0.0036	0.0051	0.0003	-0.0002	-0.0001	0.0004
H(4)	0.0010	0.0036	0.0034	0.0051	-0.0002	-0.0002	0.0003	0.0004
N(5)	-0.0032	-0.0051	-0.0010	0.0061	-	-	-	-
H(6)	0.0003	0.0012	-0.0001	0.0013	-	-	-	-
H(7)	0.0021	0.0020	0.0014	0.0032	-	-	-	-
H(8)	0.0008	0.0012	-0.0013	0.0020	-	-	-	-
N(9)	-0.0051	-0.0010	-0.0032	0.0061	-	-	-	-
H(10)	0.0020	0.0014	0.0021	0.0032	-	-	-	-
H(11)	0.0012	-0.0013	0.0008	0.0020	-	-	-	-
H(12)	0.0012	-0.0001	0.0003	0.0013	-	-	-	-
N(13)	0.0060	0.0088	-0.0065	0.0125	-	-	-	-
H(14)	-0.0014	-0.0020	0.0004	0.0025	-	-	-	-
H(15)	-0.0053	-0.0059	0.0049	0.0093	-	-	-	-
H(16)	-0.0013	-0.0008	0.0016	0.0022	-	-	-	-
N(17)	0.0088	-0.0065	0.0060	0.0125	-	-	-	-
H(18)	-0.0059	0.0049	-0.0053	0.0093	-	-	-	-
H(19)	-0.0008	0.0016	-0.0013	0.0022	-	-	-	-
H(20)	-0.0020	0.0004	-0.0014	0.0025	-	-	-	-
N(21)	-0.0010	-0.0032	-0.0051	0.0061	-	-	-	-
H(22)	-0.0013	0.0008	0.0012	0.0020	-	-	-	-
H(23)	-0.0001	0.0003	0.0012	0.0013	-	-	-	-
H(24)	0.0014	0.0021	0.0020	0.0032	-	-	-	-
N(25)	-0.0065	0.0060	0.0088	0.0125	-	-	-	-
H(26)	0.0016	-0.0013	-0.0008	0.0022	-	-	-	-
H(27)	0.0004	-0.0014	-0.0020	0.0025	-	-	-	-
H(28)	0.0049	-0.0053	-0.0059	0.0093	-	-	-	-

Table 4.6: RMSD between numerical and analytical gradient, calculated at the initial and final structure of an ac(1)-QM/MM geometry optimisation at the PBE-D3/TZVP+NPA+CGFF level of theory. Calculation of the ac-QM/MM gradient with all environment contributions (Full), with LJ parameters only (QM+LJ), with point charges only (QM+PC), and without any environment contributions(QM). All values in $E_h \cdot a_0^{-1}$.

	initial				final			
	Full	QM+LJ	QM+PC	QM	Full	QM+LJ	QM+PC	QM
benzene	0.093	0.089	0.093	0.089	0.092	0.086	0.092	0.086
urea	0.161	0.158	0.161	0.158	0.161	0.155	0.161	0.155

former are included, but the latter is neglected in the formulation of ac-QM/MM gradients. Its impact on the results is not generally negligible. It is closely related to the charge update procedure and will, thus, be discussed in the respective Section 4.6.

4.4.2 Cell Vector Gradients

Formalism of Cell Vector Gradients

One approach towards the optimisation of the cell was based on gradients in the direction of the cell vectors, referred to as the cell vector gradient (CVG) approach. As a movement of molecules had been excluded from the ac(1)-QM/MM atomic gradients by construction, the residual component of the initial QM/MM gradients may be used to describe the displacement of molecules relative to the central molecule. The approach is built upon fractional coordinates \vec{f} . These correlate to the cartesian coordinates \vec{r} as

$$\vec{r} = \underline{V}\vec{f}, \quad (4.23)$$

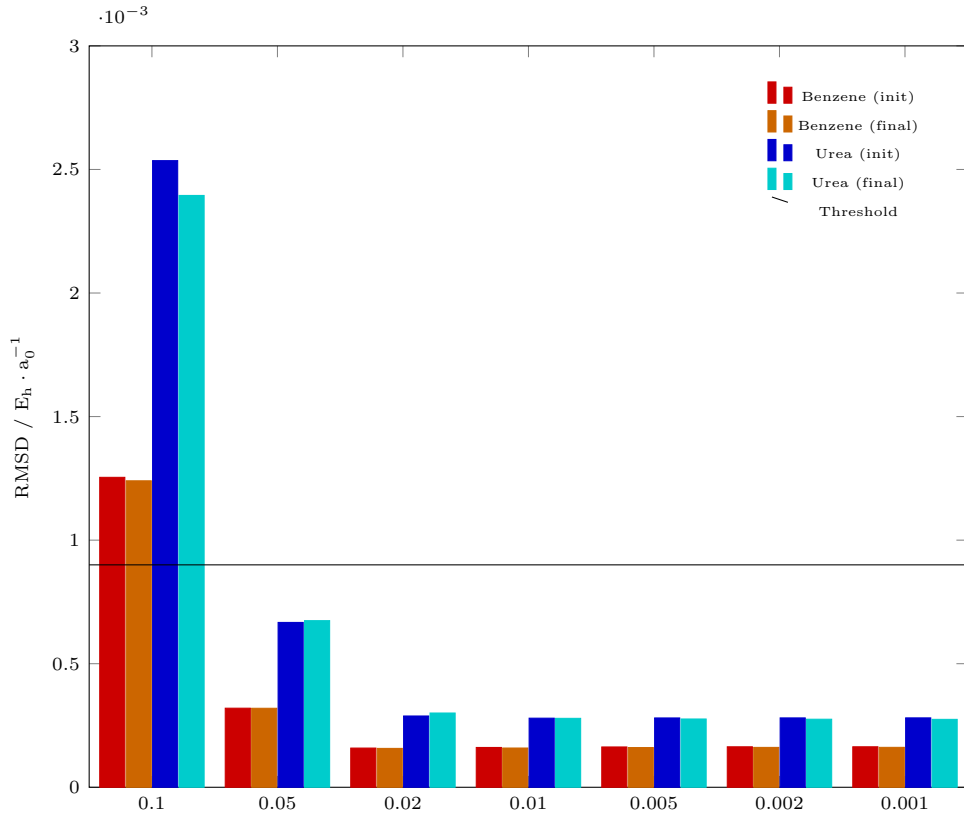


Figure 4.4: RMSD between analytical and numerical cartesian atom gradients for benzene and urea at the initial (init) and final (final) structures of ac(1)-QM/MM geometry optimisation at the PBE-D3/TZVP+NPA+CGFF level of theory. The convergence threshold (Threshold) on the RMS gradient is also indicated.

where \underline{V} is a matrix holding the cell vectors as column vectors.

This results in the formulation

$$\vec{g}_{\text{CVG}} = \sum_m \left(\left(\underline{I} \frac{1}{N} \sum_{A \in m} \vec{f}(m, A) \right) \left(\underline{V}_0 \sum_{A \in m} \nabla(m, A) \right) \right) E_{\text{cry}}. \quad (4.24)$$

There \underline{I} is the unit matrix, \underline{V}_0 consists of the normalised cell vectors as column vectors, $\vec{f}(m, A)$ and $\nabla(m, A)$ represent the respective fractional coordinate, and cartesian gradient operator of atom A in molecule m . The inner sums, similar to Equation 4.22, define a mean molecular fractional coordinate

$$\vec{f}_{\text{mol}} = \frac{1}{N} \sum_{A \in m}^N \vec{f}(m, A) \quad (4.25)$$

and the cartesian gradient for the uniform displacement of a molecule,

$$\nabla_{\text{mol}} = \sum_{A \in m}^N \nabla(m, A). \quad (4.26)$$

The multiplication $\underline{V}_0 \sum_{A \in m} \nabla(m, A)$ forms the directional derivatives for the cell vectors from the cartesian gradients. These are scaled with the mean fractional coordinates in the respective direction, yielding the expression $\underline{I} \frac{1}{N} \sum_{A \in m} \vec{f}(m, A)$. From Equation 4.23 it becomes clear that a change in the cell vectors forming \underline{V} will influence the cartesian coordinates proportionally to the fractional coordinate. For that reason, this scaling factor is required.

One obtains a gradient for the three cell vectors. This can in principle be handled like any atomic gradient. Consequently, it was chosen to implement the approach by adding a dummy atom (DA) to the coordinates.

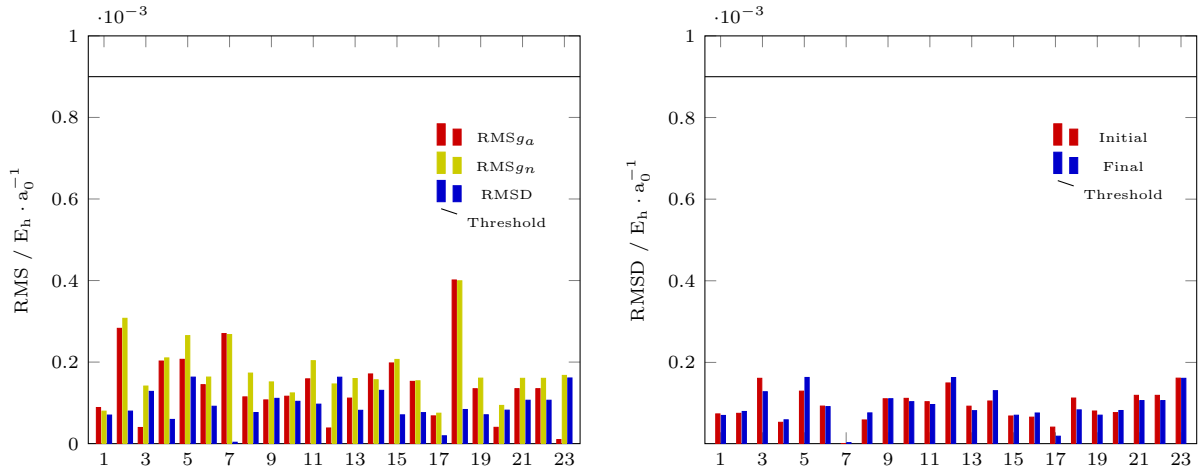


Figure 4.5: Comparison of Root Mean Square for analytical (RMS_{ga}) and numerical (RMS_{gn}) atomic cartesian gradients with the Root Mean Square Difference (RMSD) between numerical and analytical gradients for the final structure of an ac(1)-QM/MM geometry optimisation at the PBE-D3/TZVP+NPA+CGFF level of theory (left), and comparison of the final RMSD (Final) to the value obtained from the initial (Initial) structure (right). The convergence threshold (Threshold) on the RMS gradient is also indicated.

This atom is assigned the current cell vector lengths as cartesian coordinates, and the cell vector gradient as an atomic gradient. Therefore, the optimiser can treat the dummy atom like a real atom and form an optimisation step, which includes a change in the cell lengths. As cell angles are not changed, the cell vectors are simply scaled with the previous and new DA coordinates \vec{r}

$$\begin{aligned}\vec{a}^{(i)} &= \vec{a}^{(i-1)} \frac{x^{(i)}(\text{DA})}{x^{(i-1)}(\text{DA})} \\ \vec{b}^{(i)} &= \vec{b}^{(i-1)} \frac{y^{(i)}(\text{DA})}{y^{(i-1)}(\text{DA})} \\ \vec{c}^{(i)} &= \vec{c}^{(i-1)} \frac{z^{(i)}(\text{DA})}{z^{(i-1)}(\text{DA})}\end{aligned}\quad (4.27)$$

There $\vec{a}^{(i)}$ to $\vec{c}^{(i)}$ are the cell vectors and $x^{(i)}$ to $z^{(i)}$ are the coordinates of the dummy atom in step i . The matrix $\underline{\underline{V}}$ is updated accordingly and a structure change is calculated based on Equations 4.23 and 4.25, defining the atomic displacement of the i -th step as

$$\Delta\vec{r}(m, A) = \left(\underline{\underline{V}}^{(i)} - \underline{\underline{V}}^{(i-1)} \right) \vec{f}_{\text{mol}}(m) \quad \forall A \in m. \quad (4.28)$$

One inconsistency related to the CVG approach is found for cell angles different from 90°, especially in case of hexagonal structures. Molecules at equivalent positions should contribute to the gradients with different weightings. With the definition of fractional coordinates this is not the case. Six equivalent positions are marked at the edges of a 2D hexagonal structure in Figure 4.6. The molecules at (1,1) and (0,-1) should contribute to the \vec{v}_1 gradient by the same amount, as well as (0,1) and (-1,-1). However, the respective fractional coordinates are different, therefore, the contributions according to Equations 4.24 are different.

Consequently, a variant of the CVG approach, referred to as the mixed cell vector gradient approach (MCVG), was developed to overcome this counter-intuitive behaviour. It makes use of the inverse cell vectors \vec{w} , which are defined such that they, together with the cell vectors \vec{v} , fulfil the condition

$$\vec{v}_i \vec{w}_j = \delta_{ij}. \quad (4.29)$$

This corresponds to the matrix formulation

$$\underline{\underline{W}}^T \underline{\underline{V}} = \underline{\underline{I}}. \quad (4.30)$$

The inverse cell vectors are easily accessible through the inversion of $\underline{\underline{V}}$.

Although they have a different physical meaning, cell vectors and inverse cell vectors can be combined to form new coordinate systems with respective fractional coordinates. Thus, three combinations can be made

from one real cell vector \vec{v}_i and its two complementary inverse cell vectors $\vec{w}_{j,k}$ ($j, k \neq i$). The fractional coordinates obtained with the MCVG approach are shown in Figure 4.6. This does not guarantee the hexagonal symmetry, as the gradient contributions to \vec{v}_1 and \vec{v}_2 can still be different. Nonetheless, it seems to be a more balanced description in terms of equivalent molecules to have equivalent contributions.

Testing the Cell Vector Gradients

The X23 molecules were optimised at the ac(1)-QM/MM PBE-D3/TZVP+NPA+CGFF level of theory, applying both the CVG and MCVG approaches for the cell relaxation. Although all calculations converged without problems, the results (Table 4.7) show a systematic error with respect to symmetry in the case of CVG. Crystal symmetry is not enforced, except for the symmetry relation between each pair of cell molecules. All cell vector lengths are allowed to change without restrictions. Hence, lengths, which are required to be the same by symmetry, may become different during an ac-QM/MM optimisation. As an example, the two cell lengths a and b in the hexagonal structures triazine and trioxane should be the same. However, the optimised values of 9.949 and 10.102 Å (triazine) as well as 9.129 and 9.214 Å (trioxane) are notably different. While it seems plausible that due to numeric inaccuracies and an imperfect embedding slightly different parameters are obtained, the observed deviations are quite large. Moreover, for all other symmetric cases featuring only square angles, e.g. the tetragonal structure of urea, such mismatch among the equivalent parameters can not be found. In these cases much smaller deviations of significantly less than 0.001 Å are observed, compared to 0.153 and 0.085 Å for triazine and trioxane. Closer evaluation of the gradients with respect to a and b revealed that these are very similar. The percentage difference of only 0.44% (triazine) and -0.31% (trioxane) in these gradients does not explain the much larger percentage differences in the cell lengths of 1.5% (triazine) and 0.93%. Especially the differences $\Delta\Delta l$ between the length changes in a and b are not reasonable. These are calculated as

$$\Delta\Delta l = 100 \cdot \frac{\Delta b - \Delta a}{\Delta a} = 100 \cdot \frac{(a - a_0) - (b - b_0)}{a - a_0}, \quad (4.31)$$

where a_0 and b_0 are the initial, a and b the final values of the cell lengths. The values $\Delta\Delta l$ should roughly correspond to the respective gradient, but the differences of 110% (triazine) and -44% (trioxane) are by two orders of magnitude larger. One concluded that the calculation of cell vector gradients itself is correct, but the update procedure is not valid for systems that have angles different from 90° .

A distinctly different situation is observed with the MCVG approach. As can be seen from the results (Table 4.8), the large mismatch of equivalent cell parameters for the hexagonal structures triazine and trioxane is removed. There is still a difference of about 0.001 Å in both structures, but the $\Delta\Delta l$ values are now 0.36% (triazine) and -0.38% (trioxane). These fit nicely to the differences 0.44% (triazine) and -0.31% (trioxane) in the cell vector gradients. The overall results seem very reasonable, nevertheless, one

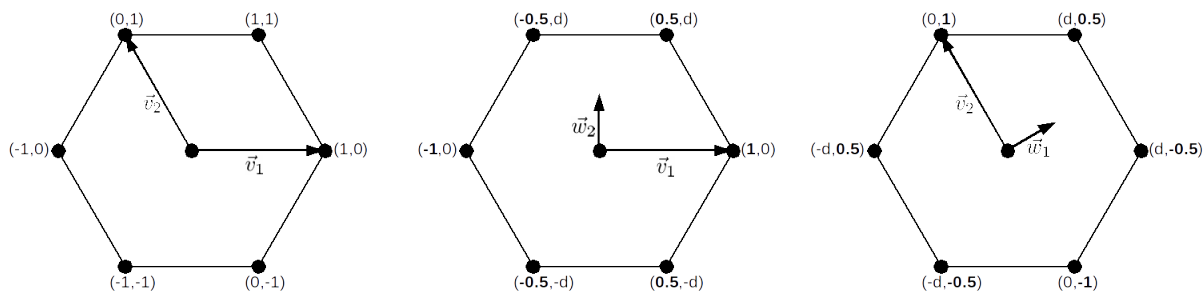


Figure 4.6: Schematic 2D hexagonal structure. The fractional coordinates of several positions are indicated according to the real cell vectors \vec{v}_1 and \vec{v}_2 (left), or according to the two combinations of one real vector \vec{v}_i and its complementary inverse vector \vec{w}_j (center and right). The respective fractional coordinates in the inverse cell vector is given in units of an arbitrary length d , which depends on the length of the corresponding real vector.

Table 4.7: Experimental (Exp.) cell vector lengths a, b, c and optimised (Opt.) values from ac(1)-QM/MM optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory, using the cell vector gradient approach. All values in Å.

	Exp.			Opt.		
	a	b	c	a	b	c
1,4-cyclohexanedione	6.650	6.210	6.870	6.616	6.237	6.804
acetic acid	13.151	3.923	5.762	13.167	3.829	5.685
adamantane	6.639	6.639	8.918	6.630	6.630	8.910
ammonia	5.1305	5.1305	5.1305	4.878	4.878	4.878
anthracene	8.414	5.990	11.095	8.412	5.970	11.088
benzene	7.390	9.420	6.810	7.356	9.384	6.771
carbon dioxide	5.624	5.624	5.624	5.429	5.429	5.429
cyanamide	6.856	6.628	9.147	6.750	6.683	8.676
cytosine	13.044	9.496	3.814	12.609	9.403	3.694
ethylcarbamate	5.051	7.011	7.543	4.970	6.983	7.457
formamide	3.604	9.041	6.994	3.380	8.678	6.364
imidazole	7.582	5.371	9.790	7.579	5.365	9.785
naphthalene	8.085	5.938	8.634	8.074	5.904	8.623
oxalic acid (α)	6.548	7.844	6.086	6.487	7.657	6.049
oxalic acid (β)	5.330	6.015	5.436	5.232	6.006	5.424
pyrazine	9.325	5.850	3.733	9.165	5.696	3.649
pyrazole	8.190	12.588	6.773	8.225	12.611	6.815
triazine	9.647	9.647	7.281	9.949	10.102	7.501
trioxane	9.320	9.320	8.196	9.129	9.214	8.014
urea	5.565	5.565	4.684	5.575	5.575	4.682

can observe convergence problems in the geometry optimisations of 4 systems (formamide, oxalic acid (β), trioxane, and urea) due to an OMR. These might relate to the step size for the cell length update. A view of the optimisation trajectories indicated that the structures were already close to convergence before the OMR occurred. Thus, it might be that the cell vector gradients are less reliable than the atomic gradients once gradients become small. Nevertheless, the results already show that reasonable cell lengths and volumes can be obtained. With the exception of pyrazole and triazine, optimised cell volumes are smaller than experimental ones, which would be expected since thermal expansion is not considered in the ac-QM/MM model. However, intermolecular interactions are solely described at the QM/MM level, wherefore a very accurate description cannot be expected. Consequently, the largest deviations of 14% and 11% are found for ammonia and urea which provide several intermolecular hydrogen bonds per molecule. Another point to be noted are the very small differences for some systems. In these cases, e.g. imidazole and naphthalene, not only the final volume, but also the cell lengths hardly change from the experimental ones. It is deemed rather unlikely to find such close agreement for several structures, but no clearly erratic behaviour of the optimisation could be determined. This might be related to the somewhat arbitrary choice of the cell vectors. In principle a larger choice of the cell would result in smaller fractional coordinates for the same cartesian position. With the gradients being independent of the choice of cell vectors, smaller cell vector gradients would result. However, the effect is not observed to be more pronounced for the large cells. We, therefore, dismiss this as a likely cause.

4.4.3 Stress and Strain

The aforementioned approaches for the cell optimisation made use of the defined cell vectors. However, these can in principle be chosen arbitrarily. Hence, it is possible, and fortunately easier, to apply a similar scheme for gradients in the three cartesian directions. In principle this leads to a microscopic analogy of the macroscopic description of stress and strain in solids.

Table 4.8: Experimental (Exp.) cell vector lengths a, b, c and volumes V , and optimised (Opt.) values from ac(1)-QM/MM optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory, using the MCVG approach. Results from optimisation runs with an OMS occurrence are included, but marked with round brackets. All lengths in Å, all volumes in Å³.

	Exp.				Opt.			
	a	b	c	V	a	b	c	V
1,4-cyclohexanedione	6.650	6.210	6.870	279.6	6.481	6.395	6.721	274.5
acetic acid	13.151	3.923	5.762	297.3	13.489	3.811	5.575	286.6
adamantane	6.639	6.639	8.918	393.1	6.562	6.562	8.680	373.7
ammonia	5.130	5.130	5.130	135.1	4.878	4.879	4.878	116.1
anthracene	8.414	5.990	11.095	456.5	8.453	5.888	11.122	451.8
benzene	7.390	9.420	6.810	474.1	7.304	9.255	6.704	453.2
carbon dioxide	5.624	5.624	5.624	177.9	5.429	5.429	5.429	160.0
cyanamide	6.856	6.628	9.147	415.6	6.750	6.758	8.721	397.8
cytosine	13.044	9.496	3.814	472.4	12.142	9.556	3.796	440.4
ethylcarbamate	5.051	7.011	7.543	248.8	4.989	7.103	7.134	235.5
formamide	3.604	9.041	6.994	224.1	(3.570)	(9.020)	(6.965)	(220.5)
imidazole	7.582	5.371	9.790	348.8	7.582	5.370	9.790	348.7
naphthalene	8.085	5.938	8.634	340.8	8.085	5.936	8.633	340.7
oxalic acid (α)	6.548	7.844	6.086	312.6	6.547	7.840	6.085	312.3
oxalic acid (β)	5.330	6.015	5.436	156.9	(5.319)	(6.014)	(5.435)	(156.5)
pyrazine	9.325	5.850	3.733	203.6	9.325	5.849	3.732	203.6
pyrazole	8.190	12.588	6.773	698.3	8.194	12.589	6.778	699.1
triazine	9.647	9.647	7.281	586.8	9.948	9.949	7.407	634.9
trioxane	9.320	9.320	8.196	616.5	(9.217)	(9.217)	(8.055)	(592.6)
urea	5.565	5.565	4.684	463.4	(5.438)	(5.438)	(4.368)	(129.2)

Derivation of Stresses from Atomic Gradients

Strains of a macroscopic system are defined as

$$\epsilon_{ij} = \frac{\Delta r_i}{r_j^0}. \quad (4.32)$$

Hereby $\Delta r_i = r_i - r_i^0$ is the deformation in the cartesian direction i , relative to the solids absolute dimension r_j in the cartesian direction j . The definition of volume normalised stresses

$$\sigma_{ij} = \frac{1}{V} \frac{\partial E}{\partial \epsilon_{ij}} \quad (4.33)$$

is in principle a generalisation of Hooke's law, introducing energy derivatives with respect to coordinates. The derivative with respect to the strain can be rewritten as

$$\frac{\partial}{\partial \epsilon_{ij}} = \frac{\partial}{\partial \frac{\Delta r_i}{r_j^0}} = r_j^0 \frac{\partial}{\partial \Delta r_i} = r_j^0 \frac{\partial}{\partial (r_i - r_i^0)} = r_j^0 \frac{\partial}{\partial r_i}, \quad (4.34)$$

since the r_i^0 are constant by definition. So, the stress components can be expressed in terms of

$$\sigma_{ij} = \frac{1}{V} r_j^0 \frac{\partial E}{\partial r_i}. \quad (4.35)$$

Although it resembles a common gradient, the coordinates r are macroscopic dimensions of a solid. Nevertheless, they may as well be interpreted in terms of molecular coordinates and their change. The relative changes should be the same from a micro- and macroscopic point of view. Thus, all molecules m need to be considered, which is a finite number in our cluster. A summation of all molecular contributions yield the stress

$$\sigma_{ij} = \frac{1}{V} \sum_m R_j^0(m) \frac{\partial E}{\partial R_i(m)}. \quad (4.36)$$

This uses the cartesian molecular coordinates

$$\vec{R}(m) = \frac{1}{N} \sum_{A \in m} \vec{r}(A), \quad (4.37)$$

calculated from the atomic coordinates $\vec{r}(A)$ to describe the displacement of the molecules as a whole. The corresponding gradients can then be obtained as

$$\frac{\partial E}{\partial R_i(m)} = \sum_{A \in m} \frac{\partial E}{\partial r_i} \quad (4.38)$$

from the sum of atomic gradient components. So, the stress components are expressed in terms of the available atomic gradients. Compacting all parts into a matrix formulation yields the final expression

$$\underline{\underline{\sigma}} = \frac{1}{V} \sum_m \left[\left(\frac{1}{N} \sum_{A \in m} \vec{r}(A) \right) \left(\sum_{A \in m} \nabla(A, m) E_{\text{cry}} \right)^T \right]. \quad (4.39)$$

The first term represents a mean cartesian coordinate of the molecule as in Equation 4.37, the second one describes a gradient for the molecular movement, i.e. of all atoms within the same molecule moving in the same direction.

Since the stress and strain tensors are symmetric, it is for some applications advantageous to use the Voigt notation, which defines

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix}^V := \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{pmatrix} \quad (4.40)$$

and

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix}^V := \begin{pmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ 2\epsilon_{23} \\ 2\epsilon_{13} \\ 2\epsilon_{12} \end{pmatrix}. \quad (4.41)$$

The calculated off-diagonal elements σ_{ij} and σ_{ji} of the stress matrix may be slightly different, so, these are averaged as

$$\begin{aligned} \sigma_4 &= \frac{1}{2}(\sigma_{23} + \sigma_{32}) \\ \sigma_5 &= \frac{1}{2}(\sigma_{13} + \sigma_{31}) \\ \sigma_6 &= \frac{1}{2}(\sigma_{12} + \sigma_{21}). \end{aligned} \quad (4.42)$$

For the optimisation a similar implementation as for the CVG approach is used. Two dummy atoms (DA) are added to the structure, with their six cartesian coordinates $r_i(\text{DA})$ holding the six strain components, and their cartesian gradient components $g_i(\text{DA})$ using the six stress components. This is not really consistent, as gradients and stresses are of different dimensions, the former having the unit of a force, the latter having the unit of a pressure. To compensate for that fact, the stress components were scaled up by a factor of S^{u}

and resulting strains were scaled down by S^d ,

$$\begin{aligned} g_i(\text{DA}) &= S^u \sigma_i \\ \epsilon_i(\text{DA}) &= \frac{1}{S^d} r_i(\text{DA}). \end{aligned} \quad (4.43)$$

For S^u a default value of about 265 was taken, which stems from the ratio of the default energy convergence criteria (in atomic units) for atomic gradients and stresses applied to plane-wave calculations in the CASTEP programme package.⁹⁹ This allows for the application of convergence thresholds in the same order of magnitude to both the atomic gradients and stresses. However, this does not guarantee that reasonable step sizes for atomic coordinates and strains are obtained.

Although not the same as gradients in a physical sense, stresses and gradients have an analogous meaning for the optimisation, hence, the stresses scaled according to Equation 4.43 are denoted as gradients. Similarly, the resulting dummy atom cartesian coordinates are accordingly called coordinates, since the strain updates have an analogous meaning to the atomic coordinate updates

$$\Delta \vec{r}(m, A) = \underline{\underline{\epsilon}} \left(\frac{1}{N} \sum_{B \in m}^N \vec{r}(m, B) \right). \quad (4.44)$$

The sum again represents a common molecular coordinate which defines the uniform displacement of all atoms within the molecule. This approach to the application of stress and strain is referred to as the external approach.

Two different internal approaches (i.e. the external optimiser is only employed for the optimisation of the intramolecular geometry) on the application of stresses make use of the elasticity tensor $\underline{\underline{C}}$. In the Voigt notation $\underline{\underline{C}}$ it is reduced to a 6×6 matrix $\underline{\underline{C}}$ with its elements defined as

$$C_{ij} = \frac{\partial \sigma_i}{\partial \epsilon_j}. \quad (4.45)$$

This leads to the relation

$$\vec{\epsilon} = \underline{\underline{C}}^{-1} \vec{\sigma}. \quad (4.46)$$

It is equivalent to the common formulation of a step in the molecular optimisation

$$\vec{r} = \underline{\underline{H}}^{-1} \vec{g} \quad (4.47)$$

with the full set of atomic coordinates \vec{r} , the Hessian matrix $\underline{\underline{H}}$, and the full set of atomic gradients \vec{g} . Hence, in principle the same methods known from molecular geometry optimisation can be used. The simplest approach is to use a unit matrix $\underline{\underline{I}}$ for $\underline{\underline{C}}$, possibly scaled with a scalar value S ,

$$\underline{\underline{C}} = S \underline{\underline{I}}, \quad (4.48)$$

which is referred to as the Steepest Descent (SD) approach.

A second, more elaborate variant was tested. In principle a numerical calculation of $\underline{\underline{C}}$ is possible, but this would at least require 21 displacement calculations for the independent elements C_{ij} . This would largely overcompensate any gain from a faster convergence behaviour. A simplified test was based on the negligence of off-diagonal elements of the elasticity tensor. This is referred to as the simplified elasticity (SE) approach. The standard formulation for numerical stresses was modified to

$$C_{ij}^{\text{approx}} = \frac{\sigma_i^{(n)} - \sigma_i^{(n-1)}}{\epsilon_j^{(n)} - \epsilon_j^{(n-1)}} \delta_{ij}. \quad (4.49)$$

So, instead of differences for the displacement of single coordinates, the differences in σ_i and ϵ_j between the

Table 4.9: Cell parameters of the 2D model structures to test the stress and strain approach. Initial values (Initial) and the optimised values (Final) obtained according to the stress and strain approach. Cell lengths a, b in Å, cell angle γ in $^\circ$, cell area A in Å².

	Initial				Final			
	a	b	γ	A	a	b	γ	A
hexagonal 1	2.117	2.117	120.0	3.880	4.951	4.951	60.0	21.231
hexagonal 2	2.117	2.117	60.0	3.880	4.951	4.951	60.0	21.231
hexagonal 3	3.666	2.117	90.0	7.760	4.951	4.951	60.0	21.231
monoclinic 1	1.673	2.117	71.6	3.360	4.951	4.951	60.0	21.231
monoclinic 2	2.245	2.117	135.0	3.360	4.951	4.951	60.0	21.231
monoclinic 3	2.182	3.086	104.0	6.721	4.951	4.951	60.0	21.231
orthogonal 1	2.117	2.117	90.0	4.480	4.887	4.887	90.0	23.884
orthogonal 1 (disp.)	2.117	2.117	88.6	4.479	4.951	4.951	60.0	21.232
orthogonal 2	2.993	2.993	90.0	8.961	4.887	4.887	90.0	23.884
orthogonal 3	4.233	4.233	90.0	17.922	4.887	4.887	90.0	23.884
orthogonal 4	2.117	2.993	120.0	4.480	4.887	4.887	90.0	23.885

current and the previous structure of the n -th optimisation step are used. In case the diagonal elements are dominant, this might be a reasonable approximation.

Simplified Test Systems

In order to evaluate the general applicability of the stress and strain approach, the latter was tested on simplified model systems. A series of 2D and 3D structures was made up of interacting atoms. Interactions were represented with arbitrarily chosen LJ parameters ($r_0 = 5.0$ Å and $\epsilon = 2.093$ kJ · mol⁻¹ (0.5 kcal · mol⁻¹)) only. Structures were generated using 2D and 3D cells with different numbers of atoms per cell and different degrees of symmetry. 2D structures were built as formal 3D structures, but with a cell parameter c much larger than the cluster radius. Thus, only atoms in the ab plane were considered, rendering the c, α, β cell parameters essentially meaningless, but allowing for the unchanged ac-QM/MM framework to be applied. For these arbitrary structures (Table 4.9) cell optimisations were carried out based on the stresses calculated from atomic gradients, and using the external approach for the cell optimisation. Starting from either a hexagonal or a quasi-monoclinic (i.e. for the determining parameters in the plane $a \neq b$ and $\gamma \neq 90^\circ$) structure and using different arbitrary choices of the unit cell (Figure 4.7) always led to the same final structure, as can be seen from Table 4.9. It is a hexagonal structure, which nicely agrees with the hexagonal closest packing expected for a 2D arrangement of spheres. Only the case of a quadratic structure did result in a relaxation of a and b parameters, but in no change of γ . However, the quadratic structure is supposed to represent a local energy extremum for which, due to its symmetry, the shearing components of the stress tensor become zero. Consequently, a small change of the initial angle γ from 90 to 88.6° resulted in the optimisation towards the aforementioned hexagonal structure. This indicates that independently of the starting structure and the definition of the cell the same energy minimum is found.

The set of 2D structures was expanded by three triclinic and one orthorhombic 3D structures with arbitrarily chosen cell parameters, see Table 4.10. Their cell parameters were optimised according to the external strain and stress approach, yielding the final parameters shown in Table 4.10. All triclinic models were found to result in a cubic densest packing (cdp). Although the initial choice of the cell does not necessarily result in the “natural” cell definition of a cdp structure (i.e. $a = b = c$ and $\alpha = \beta = \gamma = 60^\circ$), the respective cell can easily be recognised from the final structure. The cell lengths still differ by up to 0.006 Å among the test structures, and the angles deviate by up to 0.1° from the ideal values of 60°. However, it is considered very likely that these differences stem from the construction of the initial cluster with an arbitrary cell, which does not guarantee a perfectly symmetric embedding of the central molecule (i.e. a single atom) in these test structures.

Like in the quadratic 2D case, the orthorhombic starting structure does not lead to a change in its angles, thus, only relaxing the cell vector lengths and ending as a primitive cubic one. Again, a small change of the

Table 4.10: Cell parameters of the 3D model structures. Initial values (Init) and optimised values (Final) obtained following the stress and strain approach. In some cases the optimised cell vectors were redefined to obtain comparable cell parameters. Cell lengths a, b, c in Å, cell angles α, β, γ in $^\circ$, cell volumes V in Å³.

		a	b	c	α	β	γ	V
Init	triclinic 1	1.673	2.117	2.182	76.0	85.6	71.6	7.1
	triclinic 2	2.245	2.245	2.182	62.8	80.1	33.6	4.9
	triclinic 3	1.673	2.117	2.182	104.0	94.4	71.6	7.1
	orthorhombic	2.117	2.646	3.175	90.0	90.0	90.0	17.8
	orthorhombic (disp.)	2.117	2.647	3.176	87.9	89.0	88.9	17.8
Final	triclinic 1	4.857	4.861	4.858	60.0	60.0	60.0	81.1
	triclinic 2	4.858	4.860	4.857	60.0	60.0	60.0	81.1
	triclinic 3	4.857	4.861	4.858	60.0	60.0	60.0	81.1
	orthorhombic	4.758	4.760	4.761	90.0	90.0	90.0	107.8
	orthorhombic (disp.)	4.860	4.863	4.863	60.1	60.1	60.0	81.4

cell angles by about about 1-2 $^\circ$ was applied to enable the relaxation of these as well. Indeed, the following optimisation resulted in the same cdp structure that had been obtained from all triclinic 3D tests. As for the 2D cases, it is deemed very sensible to obtain a closest packing structure from an optimisation using only LJ potentials for the interactions

The excellent overall agreement of both 2D and 3D tests shows that the stress and strain approach reliably provides optimised cells and should be applicable to geometry optimisations of real systems as well.

Watch Your Step (Size)

Although the external stress approach for the cell optimisation worked well for the simplified test systems, which do not require any internal optimisation (and are based on an analytical potential), it is not clear how to handle the step sizes between the atomic gradients and the stresses. Three combinations of scaling factor S^u and S^d , according to Equation 4.43, were tested on the X23 set in order to obtain a reasonable behaviour. These different combinations are denoted as S^u/S^d , so, the simplest combination 1/1 refers to no scaling at all. The other tested combinations are 265/10, and 265/1.

In all cases the structures were optimised at the PBE-D3/TZVP+NPA+CGFF level of theory. The approaches 1/1 without any scaling, and 265/10 show an improvement in convergence, but still no sufficient reliability. For 1/1 no optimisation run reaches the 200 step limit, but 8 of 23 systems show an OMR event, which is still far from a reliable behaviour. The 265/10 calculations suffered from an OMR in 5 of 11 cases, although the problematic cases are only partially the same. Moreover, 265/10 shows a slower convergence behaviour for the successfully converging systems. The average number of steps required is 35.2, compared

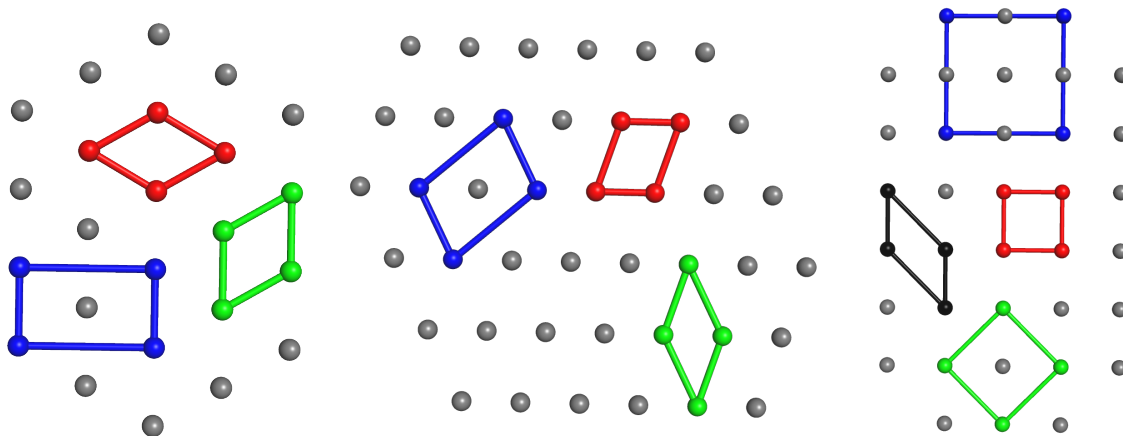


Figure 4.7: Different cell definitions for the hexagonal (left), monoclinic (middle) and orthogonal (right) 2D test systems. The respective cells 1 to 4 correspond to the colors red, green, blue, and black.

Table 4.11: Convergence behaviour of ac(1)-QM/MM cell optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory using the external stress approach. Stresses were scaled up by S^u , provided to the optimiser as gradients, and resulting strains were scaled down by $\frac{1}{S^d}$ with the combinations noted as S^u/S^d . The number of accepted optimisation steps $N(\text{acc})$ and total number of optimisation steps $N(\text{tot})$ until either convergence or the step limit of 200 is reached are given. The overall mean including all structures (Mean(Full)) and only including well converged structures (Mean(Conv)) are given as well. Values in round brackets denote calculations which are not successfully converged (due to an OMR or as the step limit is reached).

	1/1		265/10		265/1	
	$N(\text{acc})$	$N(\text{tot})$	$N(\text{acc})$	$N(\text{tot})$	$N(\text{acc})$	$N(\text{tot})$
1,4-cyclohexanedione	31	31	—	—	33	34
acetic acid	28	29	(35)	(61)	17	18
adamantane	(23)	(37)	—	—	18	18
ammonia	(13)	(29)	11	12	7	7
anthracene	30	32	—	—	38	40
benzene	8	8	—	—	8	8
carbon dioxide	4	4	6	7	8	8
cyanamide	(35)	(55)	43	60	30	30
cytosine	38	39	—	—	37	37
ethylcarbamate	48	53	—	—	42	45
formamide	17	33	28	34	17	17
hexamine	9	9	—	—	(9)	(24)
imidazole	12	12	19	28	28	29
naphthalene	32	35	—	—	16	18
oxalic acid (α)	(26)	(45)	(154)	(200)	27	28
oxalic acid (β)	25	25	(175)	(200)	30	37
pyrazine	(21)	(42)	—	—	15	16
pyrazole	24	25	(26)	(54)	23	27
succinic acid	(39)	(76)	—	—	32	37
triazine	16	22	13	16	8	8
trioxane	27	29	—	—	19	20
uracil	(54)	(80)	—	—	40	43
urea	(24)	(62)	(191)	(200)	19	21
Mean(Full)	25.4	35.3	63.7	79.3	22.7	24.8
Mean(Conv)	23.3	25.7	20.0	35.2	23.3	24.8

to 25.7 for the 1/1 calculations. Significant improvement, however, is gained from the abandonment of strain downscaling in the 265/1 case. With the only exception of hexamine all 23 optimisations converged successfully, albeit the overall convergence (24.8 steps on average) is somewhat slower than for pure internal geometry optimisation (21.9 steps on average). The single problematic system hexamine suffers from an OMR, but based on the step sizes and gradients the optimisation is considered almost converged before the OMR takes place. It seems that only the final step is affected, wherefore the results are expected to be very close to actual convergence.

Although the approaches 1/1 and 265/10 had already shown an unreliable convergence behaviour, it is still of some interest to observe the impact on (successfully or not) optimised cell parameters. Interestingly, the cell volumes for some structures change quite similarly when different scalings are applied, but not regularly. In the cases of benzene and ethylcarbamate (the only systems that converged for all scaling schemes) the volume changes with respect to the experimental values are significantly smaller for 1/1 (-0.1% and -3.5%) than for 265/1 (-4.3% and -4.8%). Nevertheless, other systems (e.g. triazine) show very comparable deviations for 1/1 and 265/1 (7.3% and 8.0%), although not all of these results are reliably converged. There seems to be no clear trend regarding the volumes, although the 1/1 scaling tends to provide rather small volume changes.

The cohesive energies show a more distinct behaviour among the scaling schemes. For almost all systems the 265/1 optimisation leads to the largest cohesive energy, thus, presumably to the closest approach to an energy minimum with respect to the cell parameters. Only for ethylcarbamate 1/1 provides a larger cohesive energy than 265/1, which might indicate a more efficient convergence behaviour of 1/1 in this case. If the OMR affected results are excluded, more than half of the systems (9 of 14) show a rather small deviation of less than $2 \text{ kJ} \cdot \text{mol}^{-1}$ in cohesive energy between 1/1 and 265/1, but larger differences up to $20.0 \text{ kJ} \cdot \text{mol}^{-1}$

Table 4.12: Cell volume deviation ΔV and cohesive energies deviation ΔE_{coh} with respect to experimental values. Results from ac(1)-QM/MM optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory using the external stress approach. Stresses were scaled up by S^u , provided to the optimiser as gradients, and resulting strains were scaled down by $\frac{1}{S^d}$ with the combinations noted as S^u/S^d . Values in round brackets denote calculations, which are not successfully converged (due to an OMR or as the step limit is reached).

	$\Delta V/\%$			$\Delta E_{\text{coh}}/\text{kJ} \cdot \text{mol}^{-1}$		
	1/1	265/10	265/1	1/1	265/10	265/1
1,4-cyclohexanedione	-1.2	–	-1.2	-5.4	–	-4.3
acetic acid	-1.1	(-1.6)	-5.0	-19.1	(-19.4)	-17.1
adamantane	(-0.8)	–	-9.0	(-0.7)	–	-1.1
ammonia	(-3.8)	-3.2	-14.4	(-2.0)	-1.6	-5.3
anthracene	-0.4	–	-1.6	-14.6	–	-14.3
benzene	-0.1	–	-4.3	-10.1	–	-11.3
carbon dioxide	-0.0	-1.4	-9.5	-7.4	-7.8	-9.4
cyanamide	(-4.7)	-0.4	-4.5	(-23.7)	-21.1	-26.8
cytosine	-5.0	–	-5.5	-55.6	–	-67.4
ethylcarbamate	-3.5	–	-4.8	-20.8	–	-16.3
formamide	-3.5	-1.0	-11.0	-48.2	-46.6	-65.4
hexamine	2.7	–	(-3.2)	-1.4	–	(-1.5)
imidazole	-0.2	-0.2	-1.7	-17.6	-17.6	-19.8
naphthalene	-0.2	–	-0.5	-11.1	–	-10.9
oxalic acid (α)	(-1.5)	(-0.3)	-3.9	(-26.6)	(-25.6)	-30.4
oxalic acid (β)	-1.4	(-1.9)	-2.0	-32.6	(-35.5)	-22.3
pyrazine	(-5.7)	–	-4.2	(-4.5)	–	-5.4
pyrazole	3.6	(2.6)	3.8	-5.5	(-5.6)	-1.1
succinic acid	(-3.4)	–	-6.7	(-50.9)	–	-52.6
triazine	7.3	6.1	8.0	14.5	14.6	15.0
trioxane	-4.2	–	-3.7	-18.8	–	-19.3
uracil	(-4.6)	–	-6.7	(-70.9)	–	-74.4
urea	(-15.7)	(-14.1)	-6.8	(-75.9)	(-66.8)	-82.3
RMSD(Full)	4.6	4.9	6.3	31.9	30.1	35.0
RMSD(Conv)	3.1	2.9	6.4	24.1	23.1	35.8

(formamide) are still found. Energies from 265/10 scaling are mostly close to the respective 1/1 energies, hence, regularly smaller than the 265/1 results.

Overall it is concluded that the 265/1 scaling approach does not only provide a more reliable convergence behaviour than the residual scaling tests, but also a very reliable convergence pattern in general. Moreover, the 265/1 scaling seems to allow for the closest approach to an energy minimum, even if only well converged results from the other scaling combinations are considered. For these reasons, the external stress approach with a 265/1 scaling was chosen as the default for the cell optimisation.

Tests on the Internal Update

Initial tests based on the steepest descent approach revealed only very small changes to the cell parameters. Therefore, increasing scaling factors S according to Equation 4.48 were used. The optimisation runs show a quite consistent behaviour with respect to the scaling for all four systems under investigation, both concerning convergence (Table 4.13), cell parameters (Table 4.14), and cohesive energies (Table 4.15). If no scaling factor is applied, convergence is achieved quickly, since almost no change in the cell parameters takes place. Nevertheless, in the case of formamide a very slow convergence is observed for the unscaled SD optimisation, showing several small updates and resulting in minimal changes from the experimental starting structure after 200 optimisation steps. This finding indicates that either the step size from an unscaled SD is too small to form a reasonable update, or the optimised cell size is very close to the initial value. However, the findings from upscaled SD optimisations strongly support the former assumption. The convergence seems to become slower for scaling factors of 10, as the number of optimisation steps until convergence increases from 14 to 20 and from 4 to 52 (ammonia and carbon dioxide, respectively). However, for cyanamide and formamide this behaviour can not be observed. All systems seem to show the fastest convergence behaviour with scaling factors of the order of 100 to 1000. However, a further increase of the scaling quickly leads to deterioration

Table 4.13: Convergence behaviour of internal steepest descent (SD) ac(1)-QM/MM cell optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory with different step size scalings. The total number of optimisation steps $N(\text{tot})$ and percentage acceptance rate of steps $P(\text{acc})$ for different SD scaling factors are shown. Results suffering from an OMR are enclosed in round brackets.

	ammonia		carbon dioxide		cyanamide		formamide	
	$N(\text{tot})$	$P(\text{acc})$	$N(\text{tot})$	$P(\text{acc})$	$N(\text{tot})$	$P(\text{acc})$	$N(\text{tot})$	$P(\text{acc})$
1	14	100	4	100	33	97	200	98
10	20	100	52	100	29	79	88	94
100	18	100	19	100	43	88	37	89
200	13	100	11	100	35	94	17	100
500	8	100	6	100	54	70	(81)	(41)
1000	7	100	10	60	(182)	(37)	(107)	(38)
2000	14	57	10	60	(182)	(33)	(112)	(39)
5000	14	57	9	56	(110)	(36)	(101)	(33)
10000	20	35	12	50	(200)	(28)	(77)	(34)
Ref.	7	100	8	100	30	100	17	100

of the results. In the cases of ammonia and carbon dioxide a 100% acceptance rate of optimisation steps drops to 35% and 50%, respectively, if the scaling factor is increased from 500 to 10000. Most likely the applied steps become too large for the larger scaling factors, hampering the convergence behaviour.

The assumption is supported by the cell volumes obtained with increasing scalings. Without scaling the optimised values remain close to the initial ones (e.g. 134.8 and 135.0 \AA^3 for ammonia), while an increase of the scaling to the order of 1000 lets the results approach the reference values from the external approach (e.g. in the case of ammonia 115.8 and 115.6 \AA^3 for a scaling factor of 500 and for the reference). Larger scaling factors do not show a monotonous change, but rather some scattering of the optimised values around the reference ones.

The impact on the cohesive energy seems to be smaller, as the energy changes with scaling is significantly smaller than the energy differences to the experimental results. Nevertheless, the same trend as for the volumes is observed for the cohesive energies. The difference with respect to the external results is in the range of 1.8 (carbon dioxide) to 17.9 $\text{kJ} \cdot \text{mol}^{-1}$ (formamide) without scaling and decreases to only 0.1 (carbon dioxide) to 3.4 $\text{kJ} \cdot \text{mol}^{-1}$ (formamide) for $S = 200$. Energy changes with increasing scaling are rather small for both ammonia and carbon dioxide which implies a rather flat energy landscape with respect to the cell parameters. This would explain the more stable convergence behaviour.

Significantly different results are obtained for cyanamide and formamide. Especially in the case of formamide the fastest convergence (17 steps) and best accordance with the external results (volume difference of 0.24% and cohesive energy difference of 3.4 $\text{kJ} \cdot \text{mol}^{-1}$, see Table 4.14) are obtained with a scaling factor of 200, but all factors ≥ 500 show an OMR and, therefore, unreliable results. A similar behaviour is found for cyanamide. In this case scaling factors of 100, 200, and 500 provide similar numbers of steps (43, 35, and 54), volume deviations (0.06, 0.04, and 0.06%), and cohesive energy deviations (2.7, 1.8, and 1.8 $\text{kJ} \cdot \text{mol}^{-1}$). However, for all larger scaling factors, an OMR leads to unclear results. Most likely the larger energy differences arising for these systems require adapted step sizes, which are not easy to obtain from the scaled SD approach. With the exception of carbon dioxide all SD results show less stabilising cohesive energies (Table 4.15), thus, suggesting that the energy minimum from the external optimisation is not exactly met. For carbon dioxide some SD energies are by about 0.1 $\text{kJ} \cdot \text{mol}^{-1}$ more stabilising than the externally optimised ones, but these small differences are deemed negligible compared to the differences of up to several $\text{kJ} \cdot \text{mol}^{-1}$ arising for the other systems. A scaling factor of 500, yielding the smallest RMSDs of 1.5 $\text{kJ} \cdot \text{mol}^{-1}$ for the energies and 0.1% for the volumes, might be the best compromise among the four test systems, but leads to convergence problems in the case of formamide. Thus, a smaller factor of 200 would seem more appropriate for the stability, but already increases RMSD values to 2.0 $\text{kJ} \cdot \text{mol}^{-1}$ and 0.3%.

Overall the application of a scaled SD approach seems possible without sacrificing accuracy, but no gain

Table 4.14: Percentage cell volume differences resulting from internal steepest descent (SD) ac(1)-QM/MM cell optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory with different step size scalings. Cell volume differences for scaling factors S from 1 to 10000 are given relative to the result of an external cell optimisation. All values in %. Results suffering from an OMR are enclosed in round brackets.

	1	10	100	200	500	1000	2000	5000	10000
ammonia	16.6	13.3	0.8	0.3	0.2	1.2	-0.1	0.8	3.4
carbon dioxide	10.5	5.9	0.8	0.4	0.0	-0.3	0.3	-1.5	-0.1
cyanamide	4.3	2.5	0.1	0.0	0.1	(-0.4)	(-0.7)	(-0.6)	(-1.8)
formamide	7.8	1.8	0.5	0.2	(0.1)	(0.1)	(-0.4)	(-0.1)	(0.4)
RMSD	10.8	7.4	0.6	0.3	(0.1)	(0.7)	(0.5)	(0.9)	(1.9)

Table 4.15: Relative cohesive energies resulting from internal steepest descent (SD) ac(1)-QM/MM cell optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory with different step size scalings. Energies for scaling factors S from 1 to 10000 are given relative to the result of an external optimisation. All values in $\text{kJ} \cdot \text{mol}^{-1}$. Results suffering from an OMR are enclosed in round brackets.

	1	10	100	200	500	1000	2000	5000	10000
ammonia	5.1	4.1	1.2	0.9	0.9	0.8	0.8	0.8	1.1
carbon dioxide	1.8	0.7	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
cyanamide	8.5	7.2	2.7	1.8	1.8	(2.8)	(3.1)	(3.2)	(3.0)
formamide	17.9	13.5	4.9	3.4	(2.2)	(1.7)	(1.7)	(1.8)	(2.1)
RMSD	10.3	7.9	2.9	2.0	(1.5)	(1.7)	(1.8)	(1.9)	(1.9)

(in case of a unified scaling even a loss) in performance is observed with respect to the external approach. Moreover, extensive tests regarding the scaling factor would be required, while convergence issues are quite possible to arise from the hydrogen bonded X23 systems. Consequently, the scaled SD approach is not investigated any further.

A more elaborate approach based on the elasticity tensor was taken into account. It was based on the neglect of off-diagonal elements of the elasticity tensor, following Equation 4.49, but initial tests already showed unreliable results for the smallest test systems. The convergence pattern for all four test systems already indicates significant errors, leading to only about 50% step acceptance for ammonia and carbon dioxide, and to an OMR for cyanamide and formamide. Moreover, for equivalent cell parameters distinctly different results are obtained, e.g. lengths of 5.050, 5.186, and 5.038 Å for the cell lengths of ammonia. Obviously the simplified elasticity tensor leads to a severe breakdown of implicit symmetry, which is deemed unacceptable. Consequently, the SE approach was not further investigated.

Both the SD and SE tests did not show any sign of superior convergence behaviour compared to the external approach. While the inclusion of more sophisticated techniques, like conjugate gradient or Broyden-Fletcher-Goldfarb-Shanno,^{100,101,102,103} could possibly improve the convergence behaviour of the cell optimisation, it was considered not very important. The possible gain from such improvements was unclear, while a very simple yet robust approach had already been established. Thus, an internal handling of the step formation for the cell parameter optimisation was not pursued.

Table 4.16: Convergence behaviour of ac(1)-QM/MM geometry optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory using different approaches to form a cell update. Either the external optimiser (External) was used, or an internal scaled simple descent approach (SD), or an internal simplified elasticity (SE) approach to form a cell update. The number of accepted optimisation steps $N(\text{acc})$ and total number of optimisation steps $N(\text{tot})$ until convergence are listed.

	External		SD		SE	
	$N(\text{acc})$	$N(\text{tot})$	$N(\text{acc})$	$N(\text{tot})$	$N(\text{acc})$	$N(\text{tot})$
ammonia	7	7	18	18	8	16
carbon dioxide	8	8	19	19	9	16
cyanamide	30	30	38	43	35	77
formamide	17	17	33	37	60	136
Mean	15.5	15.5	27.0	29.3	28.0	61.3

Table 4.17: Cell lengths a, b, c , cell angles α, β, γ , cell volumes V and cohesive energies E_{coh} from ac(1)-QM/MM optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory, using a simplified elasticity (SE) approach to generate a cell update. Lengths in Å, angles in $^\circ$, volumes in Å³, and energies in kJ · mol⁻¹.

	a	b	c	α	β	γ	V	E_{coh}
ammonia	5.050	5.186	5.038	86.4	90.2	89.4	131.7	-38.9
carbon dioxide	5.584	5.493	5.405	88.4	89.1	88.4	165.6	-36.2
cyanamide	6.713	6.684	8.936	92.7	90.9	88.9	400.4	-103.2
cytosine	3.576	8.728	6.641	92.6	96.0	89.3	205.9	-136.9

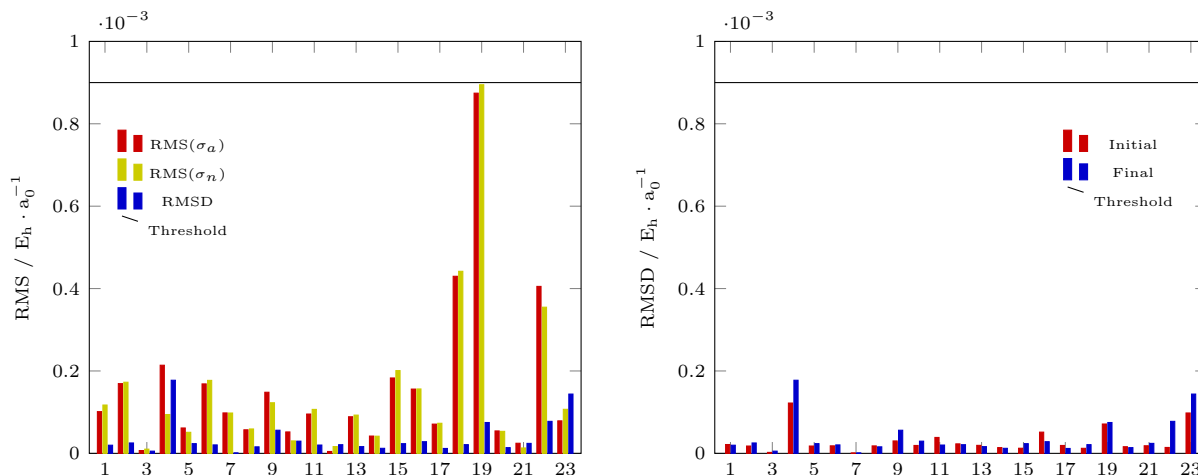


Figure 4.8: Comparison of Root Mean Square values for analytical ($\text{RMS}\sigma_a$) and numerical ($\text{RMS}\sigma_n$) stresses with the Root Mean Square Difference (RMSD) between them (left). Comparison of RMSD at the beginning (Initial) and end (Final) of an ac(1)-QM/MM geometry optimisation at the PBE-D3/TZVP+NPA+CGFF level of theory. The RMS gradient threshold (Threshold) used for the optimisation is indicated as well.

Validation of Analytical Stresses

In order to assess the general quality of the analytical stresses, these were compared to numerical stresses of the same structures. All X23 systems had been optimised at the PBE-D3/TZVP+NPA+CGFF level of theory, using analytical gradients for the internal geometry and analytical stresses for the cell optimisation. Both analytical and numerical gradients were then calculated for the initial structure and for the final optimised structure. While the stresses have a unit of $E_h \cdot a_0^{-3}$ by construction, their values are scaled and treated like gradients having a unit of $E_h \cdot a_0^{-1}$ in the optimisation process. Thus, all values provided and compared here are the modified values as they are used by the optimiser. They are used as gradients and are also given in the unit of gradients. Therefore, comparison to atomic gradient values is possible.

The results (Figure 4.8) show a similar trend as for the analytical gradients. Overall the analytical and numerical values are very similar, both at the beginning and at the end of the optimisation. For the initial structures a total RMS stress of $0.0102 E_h \cdot a_0^{-1}$ is obtained for analytical and numerical calculations, which is somewhat smaller than the RMS gradient ($0.0545 E_h \cdot a_0^{-1}$ for both analytical and numerical). The same is found for the the respective RMSD between analytical and numerical results of gradients and stresses. It is $4.1 \cdot 10^{-5} E_h \cdot a_0^{-1}$ for the stresses, comparable to $1.0 \cdot 10^{-4} E_h \cdot a_0^{-1}$ for the gradients. As expected, much smaller stresses are obtained from the optimised structures. For the whole X23 set very similar RMS stresses of $21.8 \cdot 10^{-5} E_h \cdot a_0^{-1}$ (analytical) and $22.2 \cdot 10^{-5} E_h \cdot a_0^{-1}$ (numerical) are found. The RMSD between them is of the same order of magnitude as before, namely $5.7 \cdot 10^{-5} E_h \cdot a_0^{-1}$. The behaviour is quite similar to what had been observed for the gradients.

These results indicate that the scaling factor to form a gradient from stress components is a reasonable choice and stress contributions are not heavily over- or underestimated by the optimiser. Secondly, the fact that during the optimisation stresses have dropped by the same order of magnitude as the atomic gradients, proves the external approach to be working sufficiently. Finally, the deviation between analytical and numerical stresses seems to be reasonably small to use them as a default. Therefore, analytical stresses

and the current external optimisation are applied whenever possible.

4.5 Choice of the Embedding Region

While the strongest interactions are expected to occur between the central molecule and its nearest neighbours, it is important to estimate the influence of the cluster size and the density of the background charges. The former determines the number of LJ interactions with the environment, and the latter determines the electrostatic interactions with the farther environment. For seven systems (ammonia, anthracene, benzene, oxalic acid (α), pyrazine, triazine, and trioxane) both the cutting radius $r_{\text{cluster}}/a_0 \in \{40, 50, 60, 70, 80\}$ and the number of background charges $N_{\text{bq}} \in \{108, 152, 204, 408\}$ were varied. For each combination a full optimisation of the structure was performed. The influence of both parameters on the geometry (by means of cartesian atomic coordinates, without hydrogen), cell volume and cohesive energy is determined independently. To obtain the N_{bq} dependence, for each results with $N_{\text{bq}} \neq 108$ and an arbitrary r_{cluster} , the difference to the result v (cell volume or cohesive energy) of $N_{\text{bq}} = 108$ with the same r_{cluster} is calculated,

$$\text{RMSD}(r_{\text{cluster}}) = \sqrt{\frac{1}{3} \sum_{N_{\text{bq}} \neq 108} (v(r_{\text{cluster}}, N_{\text{bq}}) - v(r_{\text{cluster}}, 108))^2}. \quad (4.50)$$

Hence, a series of differences for each r_{cluster} is obtained.

The dependence on r_{cluster} is determined equivalently. Differences from the calculation with $r_{\text{cluster}} \neq 40 a_0$ and an arbitrary N_{bq} are calculated to the results with $r_{\text{cluster}} = 40 a_0$ and the same N_{bq} .

$$\text{RMSD}(N_{\text{bq}}) = \sqrt{\frac{1}{4} \sum_{r_{\text{cluster}} \neq 40 a_0} (v(r_{\text{cluster}}, N_{\text{bq}}) - v(40 a_0, N_{\text{bq}}))^2}. \quad (4.51)$$

The results for the variation of the number of background charges (Table 4.18) and the variation of the cluster radius (Table 4.19) show an overall negligible influence of both parameters on the geometry. Only minimal changes are observed with different numbers of background charges. Independently of the cluster radius very small geometry differences (RMSDs well below 0.005 \AA), volume differences (RMSDs up to 0.1%) and cohesive energy differences (RMSDs well below $0.1 \text{ kJ} \cdot \text{mol}^{-1}$) are obtained. The influence of the N_{bq} parameter is, therefore, deemed negligible.

Larger deviations are obtained upon the variation of r_{cluster} . RMSD values for the internal geometry up to 0.005 \AA are considered negligible. In the cell volumes RMSDs up to 1.2% (oxalic acid (α)) and 0.5% (anthracene) are found with the variation of the cutoff distance. The same effect is found in the cohesive energy differences. RMSDs of up to $1.2 \text{ kJ} \cdot \text{mol}^{-1}$ (oxalic acid (α)) and $0.5 \text{ kJ} \cdot \text{mol}^{-1}$ (anthracene) result from the variation of r_{cluster} , for the residual systems the values are well below $0.2 \text{ kJ} \cdot \text{mol}^{-1}$. In the case of oxalic acid (α) several optimisations showed an unstable convergence behaviour, hence, the results are considered to possibly be artefacts. The differences for anthracene might relate to the shape and size of the molecule. Since the cluster is cut spherically around the central molecule, a large and somewhat prolate

Table 4.18: Root mean square deviation (RMSD) of internal geometry (\bar{r}) without hydrogen atoms, cell volume (V) and cohesive energy (E_{coh}) upon the variation of the number of background charges. For different cluster radii ($r_{\text{cluster}}/a_0 \in \{40, 50, 60, 70, 80\}$) the number of background charges N_{bq} was varied ($N_{\text{bq}} \in \{152, 204, 408\}$) and the deviations to the default setting $N_{\text{bq}} = 108$ were calculated. Geometry deviations in \AA , volume deviations in %, and energy deviations in $\text{kJ} \cdot \text{mol}^{-1}$.

	RMSD(40 a_0)			RMSD(50 a_0)			RMSD(60 a_0)			RMSD(70 a_0)			RMSD(80 a_0)		
	\bar{r}	V	E_{coh}	\bar{r}	V	E_{coh}	\bar{r}	V	E_{coh}	\bar{r}	V	E_{coh}	\bar{r}	V	E_{coh}
ammonia	-	0.0	0.0	-	0.0	0.0	-	0.0	0.0	-	0.0	0.0	-	0.0	0.0
anthracene	0.000	0.0	0.0	0.001	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0
benzene	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0
oxalic acid (α)	0.003	0.1	0.1	0.001	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0
pyrazine	0.000	0.1	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0
triazine	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0
trioxane	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0	0.000	0.0	0.0

Table 4.19: Root mean square deviation (RMSD) of internal geometry (\bar{r}) without hydrogen atoms, cell volume (V) and cohesive energy (E_{coh}) upon the variation of the cluster radius. For different numbers of background charges ($N_{\text{bq}} \in \{108, 152, 204, 408\}$) the cluster radius N_{bq} was varied ($r_{\text{cluster}}/a_0 \in \{50, 60, 70, 80\}$) and the deviations to the default setting $r_{\text{cluster}} = 40 a_0$ were calculated. Geometry deviations in Å, volume deviations in %, and energy deviations in $\text{kJ} \cdot \text{mol}^{-1}$.

	RMSD(108)			RMSD(152)			RMSD(204)			RMSD(408)		
	\bar{r}	V	E_{coh}	\bar{r}	V	E_{coh}	\bar{r}	V	E_{coh}	\bar{r}	V	E_{coh}
ammonia	-	0.0	0.0	-	0.0	0.0	-	0.0	0.0	-	0.0	0.0
anthracene	0.000	0.1	0.5	0.000	0.0	0.5	0.000	0.0	0.5	0.001	0.0	0.5
benzene	0.000	0.1	0.2	0.000	0.1	0.2	0.000	0.1	0.2	0.000	0.1	0.2
oxalic acid (α)	0.001	0.2	1.1	0.001	0.1	1.1	0.005	0.3	1.2	0.001	0.2	1.1
pyrazine	0.000	0.0	0.2	0.000	0.1	0.2	0.000	0.0	0.1	0.000	0.1	0.2
triazine	0.000	0.1	0.1	0.000	0.1	0.1	0.000	0.1	0.1	0.000	0.1	0.1
trioxane	0.000	0.0	0.2	0.000	0.1	0.2	0.000	0.0	0.2	0.000	0.1	0.2

molecule might not be ideally embedded in all directions. Although other structures of the X23 set are not sphere-like as well, the anthracene molecule is the largest and may serve as the best measure for the impact of the cluster size. As the RMSD value is still well below $1 \text{ kJ} \cdot \text{mol}^{-1}$ for anthracene, the impact of r_{cluster} on the results is deemed insignificant. Thus, a uniform cutoff of $40 a_0$ is used for the X23 set. Nevertheless, for larger molecules it might become necessary to enlarge the cluster size in order to ensure a reasonable embedding of the central molecule.

4.6 Charge Update Procedure

4.6.1 Implementation of the Update Scheme

General Procedure

In order to obtain consistent charges from a single molecule calculation, a charge self-consistent field (QSCF) procedure is carried out. The cluster contains charges at atom positions (atomic charges) and at the cluster surface (background charges), both providing the electrostatic embedding to be solved self-consistently.

In all cluster molecules m the equivalent atom A is assigned the same, but in principle arbitrary, charge. Based on the periodic information and these charges the background charges are fitted. The placement and fitting of these background charges is performed by a ChemShell²⁵ built-in routine. It was modified in order to restrict the fitting to the central molecule. Then atomic charges $q(\text{QM})$ are calculated from an arbitrary population analysis (by default the robust NPA scheme is employed) for the central molecule (index 0) in the embedding field. These N_a charges are compared to the current values $q(i)$. If the convergence criterion

$$\sqrt{\frac{\sum_{A \in 0}^{N_a} (q_A(\text{QM}) - q_A(i))^2}{N_a}} < q_{\text{thr}} \quad (4.52)$$

with the threshold value q_{thr} is fulfilled, the QSCF cycle is finished. Otherwise, these steps are repeated until the convergence condition is met or a maximum number of cycles has been performed. A strict threshold of 10^{-5} e was used as a default in order to obtain a tightly converged electrostatic potential.

After convergence has been achieved, the current charges are possibly scaled or damped. Then the final energy evaluation is performed. This step may be accepted or rejected by the external optimiser. In the latter case the new charges are discarded and the values from the last accepted optimisation step are restored. The general, though somewhat simplified, procedure is shown in Figure 4.9.

Introduction of Convergence Threshold and Damping

In order to improve the performance of this procedure, but also to avoid serious convergence problems of a geometry optimisation using a QSCF cycle at each optimisation step, modifications were made to the QSCF procedure. The convergence threshold q_{thr} was varied, and a damping procedure for the charges was included. It seems well justified to freeze the charges towards the end of an geometry optimisation, in order to get rid of energy contributions arising from small charge fluctuations. It is a similar approach as the freezing of the integration grid in DFT once energy changes become small. A larger threshold q_{thr} could already solve the problem, but might also hamper the consistency and introduce an error, as it is uniformly applied throughout the whole geometry optimisation. Hence, a straightforward approach to achieve a smooth transition from the unchanged update to the freezing regime is the application of a damping scheme. This provides a new charge

$$q_A(i) = s_{\text{damp}} q_A(i-1) + (1 - s_{\text{damp}}) q_A(\text{QM}) \quad (4.53)$$

of the i -th geometry optimisation step, where $s_{\text{damp}} \in [0, 1]$ is a damping factor, $q_A(i-1)$ is the previous charge of the respective atom A , and $q_A(\text{QM})$ is the newly calculated QM charge of this atom. Three approaches are used to determine the damping factor. The simplest one takes a predefined, constant value which is used throughout the whole optimisation. A second implementation makes use of the step acceptance information. If a geometry optimisation step is rejected for whatever reason, s_{damp} is increased. Otherwise, the damping factor is reset to its initial value $s_{\text{damp}}^{(0)}$,

$$s_{\text{damp}}^{(i)} = \begin{cases} s_{\text{damp}}^{(0)} & \text{if the last optimisation step has been accepted} \\ \min(s_{\text{damp}}^{(i-1)} + 0.1, 1.0) & \text{else.} \end{cases} \quad (4.54)$$

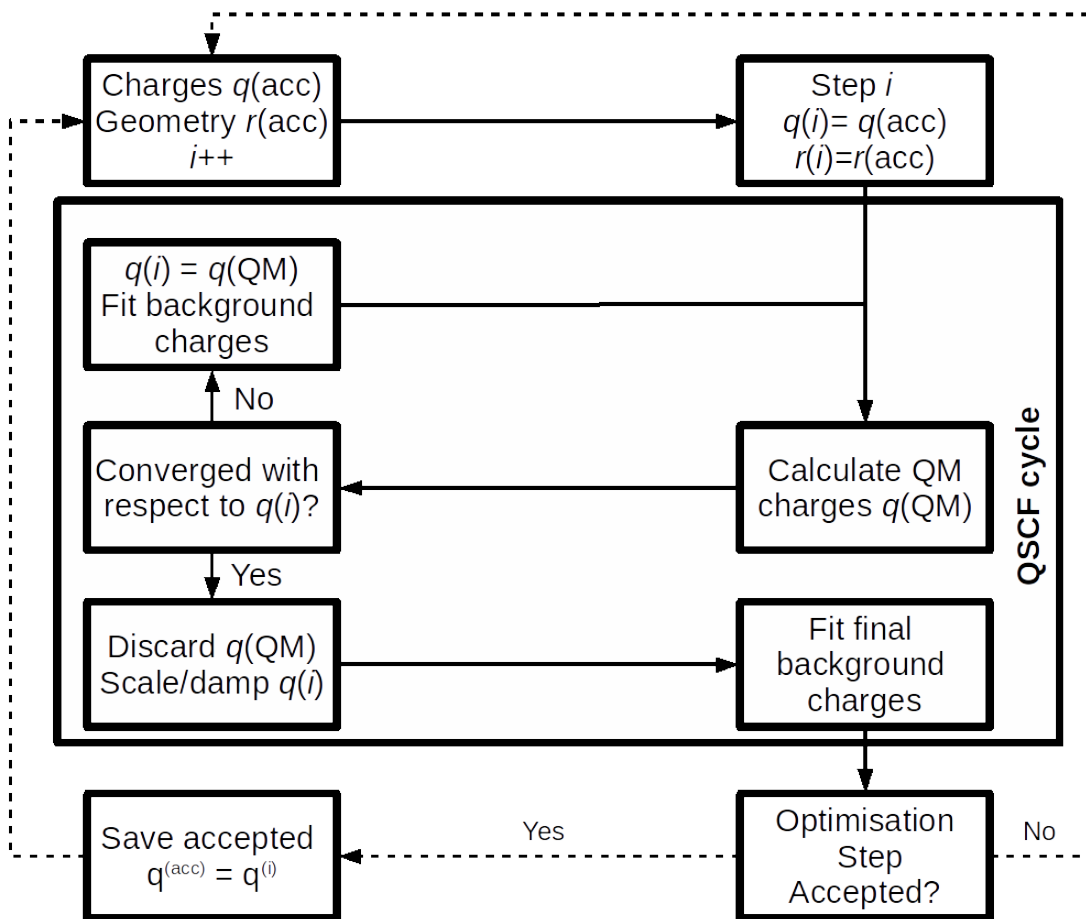


Figure 4.9: Schematic representation of the QSCF cycle during a geometry optimisation. The dashed lines represent the residual programme flow outside the QSCF cycle.

However, the largest influence of the damping would be expected to arise at the end of a geometry optimisation, as energy differences become small. Hence, s_{damp} was varied as a function of the energy difference ΔE between the last accepted optimisation steps. In order to achieve a smooth transition between no damping for large ΔE (above an upper threshold ΔE_u) and full damping for small ΔE (below a lower threshold ΔE_l), the damping function

$$s_{\text{damp}}^{(i+1)}(\Delta E) = \begin{cases} 0 & |\Delta E| > \Delta E_u \\ \frac{\ln(|\Delta E|) - \ln(\Delta E_u)}{\ln(\Delta E_l) - \ln(\Delta E_u)} & \Delta E_l \leq |\Delta E| \leq \Delta E_u \\ 1 & |\Delta E| < \Delta E_l \end{cases} \quad (4.55)$$

is used. There $|\Delta E| = |E^{(i)} - E^{(\text{acc})}|$ is the energy difference between the current optimisation step and the last accepted optimisation step, so the factor is always set at the end of one step i for the next step $i + 1$.

The latter approaches can also be combined, thus, a damping factor may be set up according to Equation 4.55 and additionally be increased for rejected steps according to Equation 4.54. Both can be done in conjunction with different q_{thr} values.

A side note to the whole QSCF procedure concerning the background charges has to be made. These are fitted using a ChemShell routine, but the number of background charges (108 by default) is larger than the number of parameters to be fitted to, namely $4N_a$ (3 electric field and 1 electrostatic potential values at each of N_a QM atoms), at least for the X23 set with a maximum number of atoms N_a per molecule of 26. This represents an overfitting problem, therefore, minimal changes in the atomic charges may severely change the resulting background charges. Hence, it needs to be evaluated whether the resulting background

Table 4.20: Convergence behaviour of unrestricted ac(1)-QM/MM geometry optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory with different values of the convergence threshold q_{thr} (in e) for the QSCF cycle. The number of optimisation steps until convergence is shown for an ac(1)-QM/MM PBE-D3/TZVP+Pop+CGFF (Pop=NPA,Chelpg) optimisation. The maximum number of optimisation steps was set to 200, thus, optimisation runs reaching this limit are expected to be not converging. Results from calculations reaching convergence only after an OMR are enclosed with round brackets.

	NPA				Chelpg			
	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
acetic acid	19	(79)	200	200	28	25	200	200
ammonia	7	9	14	13	16	200	200	200
carbon dioxide	8	8	8	8	9	9	12	13
cyanamide	32	44	200	200	(65)	200	200	200
formamide	16	25	200	200	34	25	200	200
imidazole	29	200	200	200	31	200	200	200
oxalic acid (α)	27	29	(84)	200	23	200	200	200
oxalic acid (β)	(59)	200	200	200	200	200	200	200
pyrazole	27	28	24	28	18	(116)	51	200
triazine	8	8	8	10	9	9	11	14
urea	21	200	200	200	20	200	200	200
Mean(Tot)	23.0	75.5	121.6	132.6	41.2	125.8	152.2	166.1
Mean(Conv)	19.4	21.6	13.5	14.8	20.9	17.0	24.7	13.5

charges perform reasonably.

4.6.2 Convergence Threshold

Geometry Optimisation Performance

In order to study the impact of the convergence threshold q_{thr} , a subset of eleven molecules (S11) from the X23 set was used. For all structures q_{thr} was varied from 10^{-5} e up to 10^{-2} e. Additionally, calculations were performed without any charge update (NoUp). A full geometry optimisation at the PBE-D3/TZVP level of theory with CGFF LJ parameters and either NPA or Chelpg charges was performed for each q_{thr} value to observe the convergence behaviour.

The results in Table 4.20 show two general trends for the population schemes employed. In comparison, NPA tends to allow for convergence with a looser criterion q_{thr} than Chelpg. However, even with the least restrict convergence of $q_{\text{thr}} = 10^{-2}$ e, not all structures could be converged successfully. With the latter threshold, for NPA oxalic acid (β), and for Chelpg oxalic acid (β) and cyanamide remained unconverged. For the successfully converged optimisations no advantageous behaviour of either NPA or Chelpg is observed. Only with the loosest criterion of $q_{\text{thr}} = 10^{-2}$ several structures can be compared, and the average number of optimisation steps required are very close, 19.4 with NPA and 20.9 with Chelpg.

Secondly, it can be recognized that only hydrogen bonded systems are troublesome with respect to the convergence. The systems without hydrogen bonds (carbon dioxide and triazine within S11) did not cause any problems with NPA and Chelpg, most of the others did for both of them. However, ammonia and pyrazole could be converged well with NPA even under the most rigid criterion $q_{\text{thr}} = 10^{-5}$ e. This finding correlates with the intermolecular hydrogen bond lengths (Table 4.21). The longest (2.127 and 1.834 Å with NPA) are found for these two systems, while the most difficult case, oxalic acid (α), is among the shortest lengths. With Chelpg charges the most complicated cases oxalic acid (β) and cyanamide are still among

Table 4.21: Shortest intermolecular hydrogen bond length (in Å) for the S11 structures. Values are taken from the last accepted structure of an ac(1)-QM/MM optimisations at the PBE-D3/TZVP+Pop+CGFF (Pop=NPA,Chelpg) level of theory with $q_{\text{thr}} = 10^{-5}$ e.

	acetic acid	ammonia	cyanamide	formamide	imidazole	oxalic acid (α)	oxalic acid (β)	pyrazole	urea
NPA	1.541	2.127	1.800	1.572	1.738	1.627	1.595	1.834	1.649
Chelpg	1.496	2.029	1.595	1.556	1.686	1.578	1.577	1.823	1.544

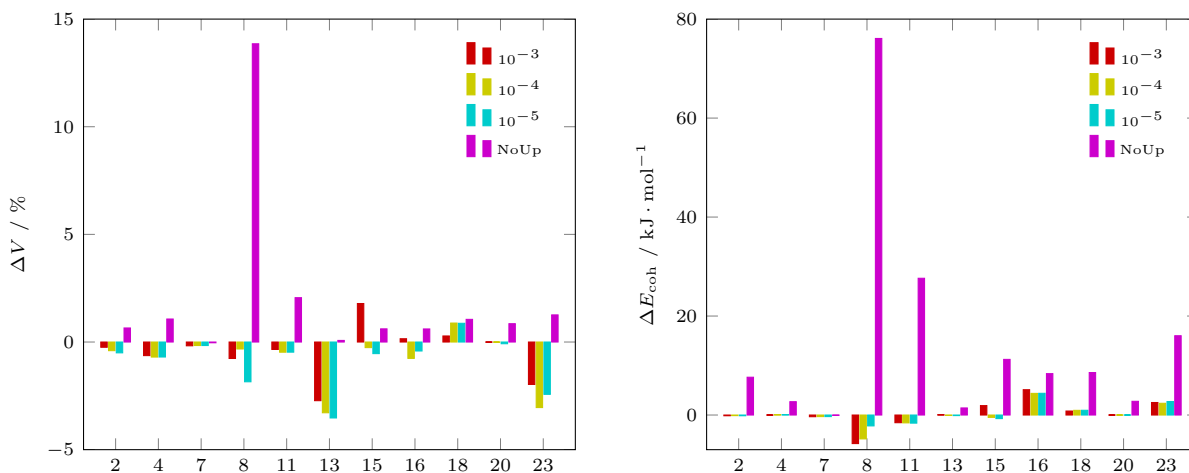


Figure 4.10: Cell volume deviation ΔV and cohesive energy deviation ΔE_{coh} for structures optimised with different QSCF convergence criteria q_{thr} (in e) or without any charge update (NoUp). Calculations at the ac(1)-QM/MM PBE-D3/TZVP+NPA+CGFF level of theory. Differences are calculated with respect to results obtained for $q_{\text{thr}} = 10^{-2}$ e. Not fully converged results are included.

the short ones, but the effect is less distinct. Nevertheless, it supports the assumption that the observed convergence problems are connected to charges in proximity to the central molecule.

Although the convergence behaviour becomes more reasonable with larger threshold values, it is not clear how much the results depend on this parameter. Since most of the structures could be converged well with the largest threshold of $q_{\text{thr}} = 10^{-2}$ e, these results were used as reference values to compare to. Additionally, comparison to results obtained without any charge update was made. In this case the atomic charges were calculated and converged for the initial geometry, but never changed during the optimisation. Background charges, however, were recalculated. As these account for the bulk electrostatic potential, the change of molecular geometry requires a change of this potential, even if the atomic charges remain unchanged.

The deviation in cell volume and cohesive energy with respect to the reference results are shown in Figures 4.10 (with NPA) and 4.11 (with Chelpg). Albeit a significant number of structures from unconverged optimisations is included, it can clearly be observed that overall the largest deviations are observed for calculations without the atomic charge update. For the cell volumes there are some large volume differences of more than 2% with smaller threshold values, e.g. for imidazole (#13) and urea (#23). Much more pronounced is the difference for E_{coh} . All results with $q_{\text{thr}} < 10^{-2}$ e, even if the optimisation has not converged successfully, are within about $6 \text{ kJ} \cdot \text{mol}^{-1}$ from the results with $q_{\text{thr}} = 10^{-2}$ e. However, the optimisations without update provide several differences of more than $5 \text{ kJ} \cdot \text{mol}^{-1}$, up to almost $80 \text{ kJ} \cdot \text{mol}^{-1}$ (cyanamide). These observations are much more distinct once only the well converged results are taken into consideration for $q_{\text{thr}} < 10^{-2}$ e. With the exception of oxalic acid (α) (#15) all volume differences are well below 1% and tend to be smaller than without the charge update, although that is not very clear due to the small number of fully converged calculations. The same is found for the cohesive energies, showing a maximum deviation of well below $1 \text{ kJ} \cdot \text{mol}^{-1}$ for any q_{thr} , with one outlier (cyanamide, (#8)).

Qualitatively, very similar results are obtained with the use of Chelpg charges instead of NPA. The overall volumes are significantly larger, the cohesive energies show slightly larger deviations than in the case of NPA. However, since optimisation runs with Chelpg charges were found to suffer from convergence problems more often, it is likely a result of the convergence pattern. With Chelpg the limitation to properly converged structures (Figure 4.11) does exclude most of the calculations for comparison, but it is recognisable that the results for carbon dioxide (#7) and triazine (#20) are very similar for both population schemes. Albeit this is not a too meaningful result, it indicates that both NPA and Chelpg tend to behave quite similarly for the benchmark structures. However, the better convergence behaviour of NPA charges sets the latter as the best choice for the ac-QM/MM method.

From the test calculations it becomes clear that for both NPA and Chelpg charges a rather loose general

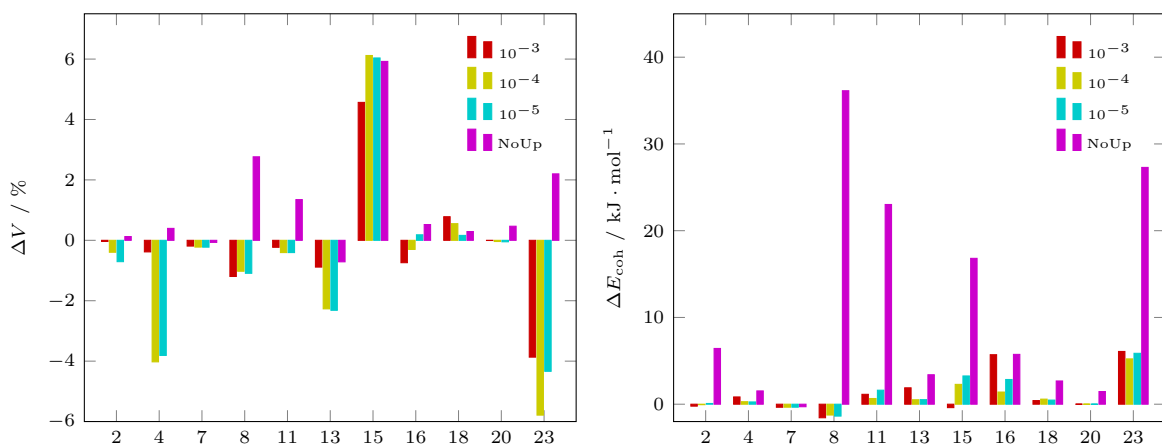


Figure 4.11: Cell volume deviation ΔV and cohesive energy deviation ΔE_{coh} for structures optimised with different QSCF convergence criteria q_{thr} (in e) or without any update (NoUp). Calculations at the ac(1)-QM/MM PBE-D3/TZVP+Chelpg+CGFF level of theory. Differences are calculated with respect to results obtained for $q_{\text{thr}} = 10^{-2}$ e. Not fully converged results are included.

threshold of $q_{\text{thr}} = 10^{-2}$ e should be considered as the default value. Nonetheless, the charge update procedure needs some additional modification to achieve a stable convergence behaviour for the whole test set.

Monitoring Geometry Optimisation Parameters

To gain some insight into this problem, the change of charges, energy differences, and gradient and step sizes were monitored during optimisations with the most rigid threshold $q_{\text{thr}} = 10^{-5}$ e. Figure 4.12 shows the change of atomic charges during up to 50 steps of geometry optimisation of the systems carbon dioxide and cyanamide. These represent the sample performance of well-behaved (carbon dioxide) and problematic (cyanamide) systems observed throughout the benchmark set. In some cases (e.g. carbon dioxide) both the charges and the geometry optimisation converge smoothly within a couple of steps. Nevertheless, especially the hydrogen bonded structures (e.g. cyanamide) tend to show a more diverging behaviour. The RMSDs between new and old charges during the first optimisation steps are by an order of magnitude larger for cyanamide than for carbon dioxide. The RMSD converges slower and tends to significantly increase in between. This happens after about 10 and 15 steps for cyanamide with Chelpg and NPA charges, respectively. Although an overall converging behaviour of the charges is observed, the corresponding geometry optimisations do not show a stable convergence profile.

While energy differences and RMS step lengths seem to decrease towards convergence, the RMS gradient (Figure 4.12) stays well above the threshold and prevents a convergence of the optimisation. The behaviour of RMS gradients and energy differences represent a quasi-breakdown of the geometry optimisation after about 15-20 steps. As can be observed from the geometry optimisation pathways as well, the steps become very small, but the convergence criteria are not met.

This erratic convergence behaviour is most likely a result of the charge update after each optimisation step. Gradients do account for the change of charge positions, but not for the change of charge values. As a consequence, an optimisation step is formed, leading to a lower energy with the previous charges. After the charge update, however, the new energy may be above the previous one and the step is rejected. In several cases, mostly the hydrogen bonded systems like cyanamide, optimisation steps are rejected repeatedly, which leads to decreasing step sizes with two possible outcomes. Either very small steps are accepted, in this case the optimisation progress becomes too slow to be finished in a reasonable number of steps. Or the steps become extremely small, leading to an OMR, as described in Section 3.1. Unfortunately, after an OMR the optimiser's internal reference geometry and energy seem to be changed. Therefore, the dl-find optimiser may accept an optimisation step, since its energy is lower than the dl-find reference energy. However, the ac-QM/MM routine can reject the same step, as it uses a different reference energy. In that case, the geometry

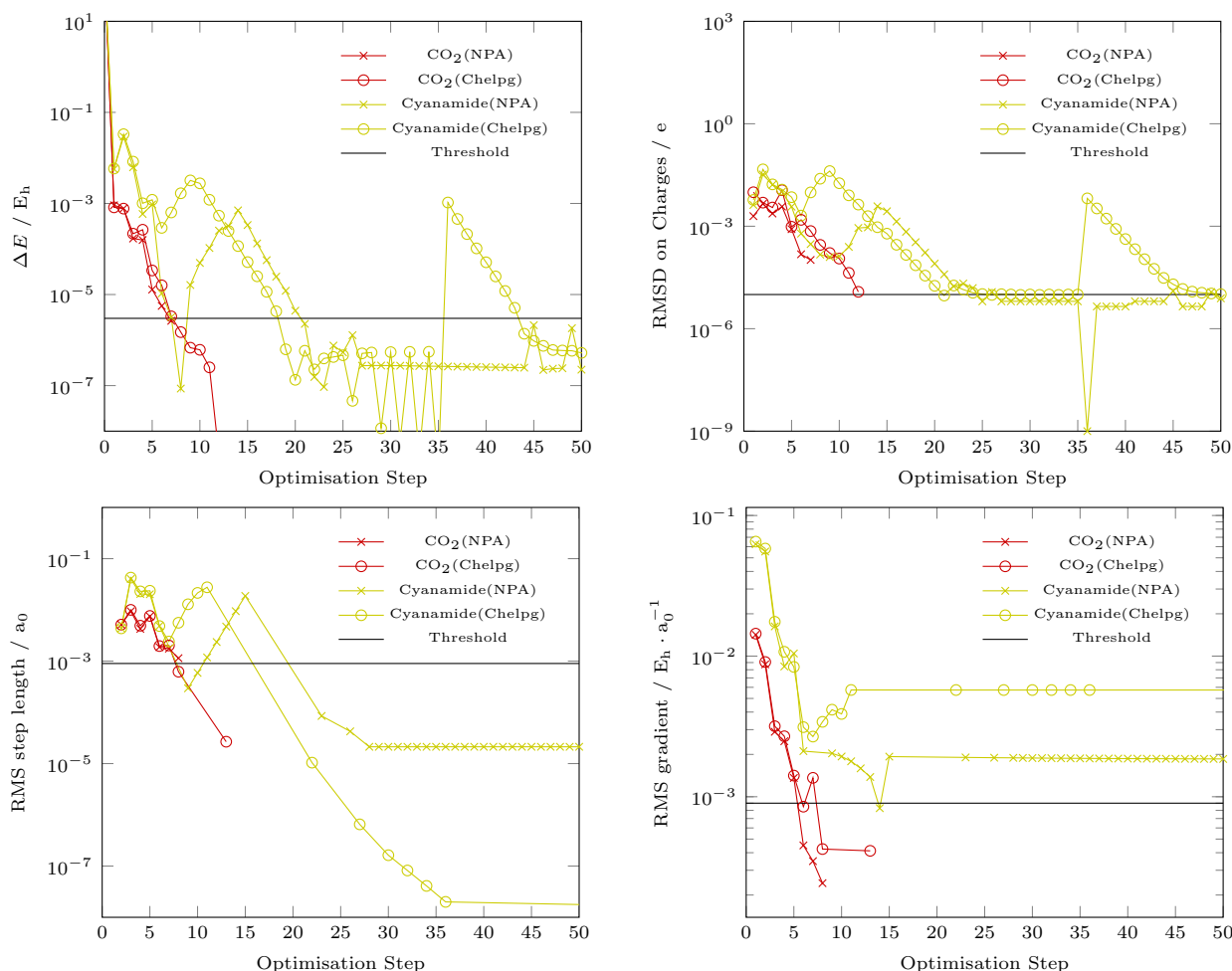


Figure 4.12: Energy difference (top left), RMSD of atomic charges (top right), RMS step length (bottom left), and RMS gradient (bottom right) with respect to the last accepted optimisation step. Values from an ac(1)-QM/MM geometry optimisation at the PBE-D3/TZVP level of theory with CGFF LJ parameters and either NPA or Chelpg charges. Thresholds of the parameters are also shown (Threshold). A rigid threshold $q_{\text{thr}} = 10^{-5} e$ was applied to the calculations.

update of the environment is discarded. As a consequence, the central molecule will change differently than the environment. Moreover, the reference geometry is changed. The optimiser uses changed reference coordinates $\vec{R}_{\text{ref,dlf}}$, and applies a geometry update $\Delta\vec{R}_{\text{dlf}}$ to calculate new coordinates

$$\vec{R}_{\text{dlf}} = \vec{R}_{\text{ref,dlf}} + \Delta\vec{R}_{\text{dlf}}. \quad (4.56)$$

The ac-QM/MM module has reference coordinates $\vec{R}_{\text{ref,ac}}$. Then the geometry update is applied as

$$\vec{R}_{\text{ac}} = \vec{R}_{\text{ref,ac}} + \Delta\vec{R}_{\text{dlf}} \quad (4.57)$$

and transferred to the cluster molecules accordingly. For the next step, energy and gradients are calculated at the geometry \vec{R}_{ac} , but interpreted by dl-find as values for the geometry \vec{R}_{dlf} . This erratic behaviour leads to sudden peaks in the energy, which can be seen in Figure 4.12.

4.6.3 Charge Damping

All subset structures were optimised at the PBE-D3/TZVP+Pop+CGFF (Pop=NPA,Chelpg) level of theory with different damping procedures. Constant damping factors up to $s_{\text{damp}} = 1.0$, which corresponds to no update of the initial atomic charges at all were applied, as well as variable damping factors. Following Equation 4.54, a damping factor increasing from 0.0 (referred to as increasing damping, IncDmp) was used.

Table 4.22: Convergence behaviour of unrestricted ac(1)-QM/MM geometry optimisation at the PBE-D3/TZVP+NPA+CGFF level of theory with different values of the charge damping factor s_{damp} . Either a constant number, an increasing value (IncDmp) or a dynamic value (DynDmp) was used. The number of optimisation steps until convergence is shown. The maximum number of optimisation steps was set to 200, thus, optimisation runs reaching this limit are not converged. Results from calculations reaching convergence only after an OMR are enclosed with round brackets. Mean values including all calculation (Mean(Tot)) and only the well converged results (Mean(Conv)) are given.

	0.0	0.25	0.5	0.75	1.0	IncDmp	DynDmp
acetic acid	200	200	23	26	17	(49)	18
ammonia	13	10	15	20	7	11	7
carbon dioxide	8	8	9	12	8	8	8
cyanamide	200	200	(57)	45	44	(48)	32
formamide	200	200	17	(50)	17	(144)	16
imidazole	200	200	13	(45)	32	(60)	30
oxalic acid (α)	200	(63)	200	28	26	(144)	26
oxalic acid (β)	200	200	(32)	(47)	27	200	200
pyrazole	28	200	(42)	(43)	26	25	27
triazine	10	9	13	21	8	10	8
urea	200	200	24	31	17	200	17
Mean(Tot)	132.6	135.5	40.5	33.5	20.8	81.7	35.4
Mean(Conv)	14.8	9.0	16.3	26.1	20.8	13.5	18.9

Secondly, a dynamic damping factor following Equation 4.55 was applied in conjunction with the increase upon rejected steps according to Equation 4.54 (referred to as dynamic damping, DynDmp). If no damping procedure is applied, the calculation is labelled as NoDmp. Calculations with constant $s_{\text{damp}} = 1$, thus, using an initially calculated set of charges during the whole optimisation (NoUp), were included to estimate the relevance of charge updates for the results.

Constant Damping

The results in Tables 4.22 and 4.23 show the same trend for NPA and Chelpg charges. Obviously all calculations using constant charges during the geometry optimisation converge smoothly, which confirms the assumption that, indeed, the charge update is responsible for the convergence issues. Albeit a general improvement of the convergence can be observed for larger damping factors, there seems to be no clear threshold for an overall convergence. For formamide (#11) and imidazole (#13) with NPA, and for oxalic acid (β) (#16) with Chelpg, an oscillation is found. The respective calculations with $s_{\text{damp}} = 0.5$ converge, but with both smaller $s_{\text{damp}} = 0.25$ and larger $s_{\text{damp}} = 0.75$ no convergence is achieved. In case of NPA charges, for oxalic acid (β) (#16) and pyrazole (#18) none of the damping factors enables convergence of the geometry optimisation. In total, 7 (NPA) and 10 (Chelpg) of the 11 structures could be converged with $s_{\text{damp}} = 0.75$, 7 (NPA) and 8 (Chelpg) of 11 could be converged with $s_{\text{damp}} = 0.5$. Although it is an improvement with respect to 4 (NPA) and 2 (Chelpg) converged structures for the undamped case, it is definitely not sufficient to be regularly applied. Moreover, the required number of optimisation steps for the well converged cases increases with the damping factor, unless the update is turned off completely with $s_{\text{damp}} = 1.0$. With NPA an increase from 14.8 to 26.1 steps is obtained when s_{damp} changes from 0.0 to 0.75, with Chelpg the increase is from 13.5 to 33.2 steps. Although these numbers are of limited validity, as only a limited number of the calculations is well converged, the same trend is observable for individual systems, e.g. for carbon dioxide the increase is from 8 to 12 steps with NPA and from 13 to 18 steps with Chelpg. Thus, it seems that on the one hand the robustness is improved with an increased damping, but on the other hand optimisations are notably slowed down.

Adaptive Damping

No improvement is gained from the simple increase of damping. At first glance the convergence seems to be somewhat better than without damping. However, in several cases an OMR is found (as described above), thus, the faster convergence is likely just an artefact of the memory reset. The damping might improve

Table 4.23: Convergence behaviour of unrestricted ac(1)-QM/MM geometry optimisation at PBE-D3/TZVP+Chelpg+CGFF level of theory with different values of the charge damping factor s_{damp} . Either a constant number, an increasing value (IncDmp) or a dynamic value (DynDmp) was used. The number of optimisation steps until convergence is shown. The maximum number of optimisation steps was set to 200, thus, optimisation runs reaching this limit are not converged. Results from calculations reaching convergence only after an OMR are enclosed with round brackets. Mean values including all calculation (Mean(Tot)) and only the well converged results (Mean(Conv)) are given.

	0.0	0.25	0.5	0.75	1.0	IncDmp	DynDmp
acetic acid	200	23	15	36	20	(41)	24
ammonia	200	200	15	26	16	(80)	14
carbon dioxide	13	10	10	18	8	13	9
cyanamide	200	200	35	50	26	200	30
formamide	200	16	23	40	15	(112)	36
imidazole	200	200	200	45	14	(61)	34
oxalic acid (α)	200	200	200	44	25	200	26
oxalic acid (β)	200	200	23	(98)	36	(82)	39
pyrazole	200	56	38	19	18	200	20
triazine	14	9	12	19	8	14	8
urea	200	21	26	35	17	200	20
Mean(Tot)	166.1	103.2	54.3	39.1	18.5	109.4	23.6
Mean(Conv)	13.5	22.5	21.9	33.2	18.5	13.5	23.6

convergence, so, the calculations could be close to convergence when the OMR takes place. However, there is no solid indication that the increasing damping results are closer to convergence than those without damping.

A much larger impact on the convergence can be observed if the damping factor as a function of the energy difference is used. In these cases a generally smooth convergence is observed for all systems with both NPA and Chelpg. The only exception is oxalic acid (α) (#15) with NPA charges, the respective optimisation did not converge within 200 steps. Nevertheless, since 94% of all optimisation steps are accepted, it is assumed that the slow convergence is rather caused by the optimiser and not related to the charge update. The overall convergence with this damping is even somewhat faster than in case of no charge update (18.9 compared to 20.8 steps) with NPA, but slower (23.6 to 18.5 steps) with Chelpg. Therefore, no dramatic loss in performance is expected from this damping approach.

Unfortunately, a stable convergence behaviour is no guarantee for reasonable results. Hence, one needs to estimate the impact of the dynamic damping on the final volumes and energies.

Comparison of Constant Damping and Adaptive Damping Results

In two cases a reliable convergence behaviour was observed, namely for the dynamic damping and for the constant damping with $s_{\text{damp}} = 1.0$. The dynamic damping calculations are estimated to be more reasonable, because they include a charge update. Hence, the dynamic damping case is used as a reference to compare the results of other damping approaches. Albeit several of those results are not converged, they may still serve as an estimate of the difference to be expected among the different damping approaches.

The final cell volume and cohesive energy differences in Figures 4.13 and 4.14 show two significant trends. The largest deviations are found for $s_{\text{damp}} = 1.0$ calculations, especially cohesive energies show deviations up to $77.0 \text{ kJ} \cdot \text{mol}^{-1}$ (cyanamide with NPA). All optimisations using any damping show much smaller absolute deviations of up to $5.5 \text{ kJ} \cdot \text{mol}^{-1}$ (cyanamide), although the unconverged optimisations were included. For NPA the RMSD values with respect to the reference energies are well below $2 \text{ kJ} \cdot \text{mol}^{-1}$ for $s_{\text{damp}} < 1.0$, but much larger ($25.6 \text{ kJ} \cdot \text{mol}^{-1}$) for $s_{\text{damp}} = 1.0$. The same behaviour is observed for Chelpg charges, with RMSDs well below $3 \text{ kJ} \cdot \text{mol}^{-1}$ for $s_{\text{damp}} < 1.0$, but $16.3 \text{ kJ} \cdot \text{mol}^{-1}$ for $s_{\text{damp}} = 1.0$. Many energies for $s_{\text{damp}} \leq 1.0$ are slightly lower than the reference, but not necessarily well converged. Therefore, one concludes that the dynamic damping reference energies tend to be slightly too high. Nevertheless, these differences are negligibly small compared to the differences arising from the neglect of the charge update. Hence, one does not expect large energy errors for the dynamic damping scheme.

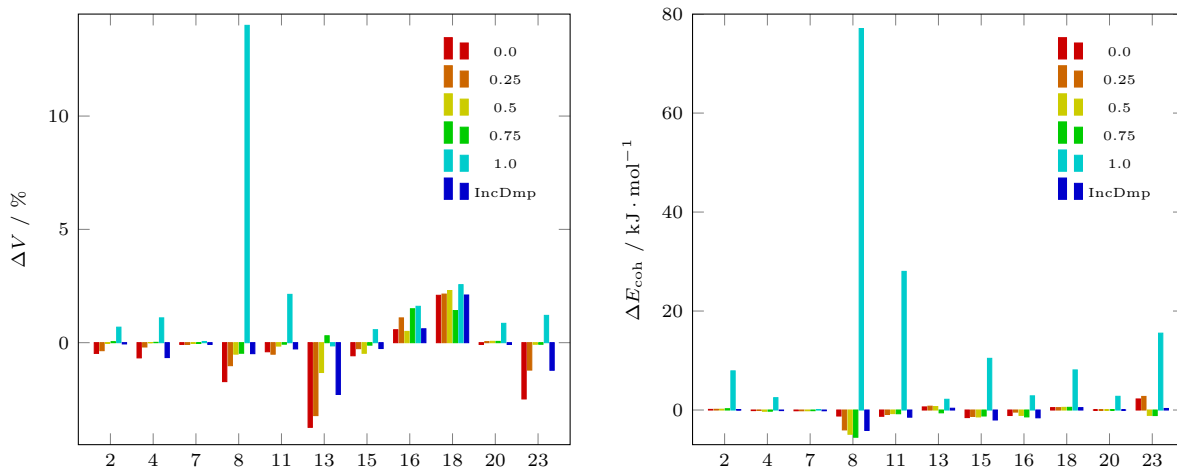


Figure 4.13: Cohesive energy differences ΔE_{coh} and cell volume differences ΔV for different damping factors s_{damp} , always calculated with respect to results from analogous calculations using dynamic damping. Results from ac(1)-QM/MM optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory. Not fully converged results are included.

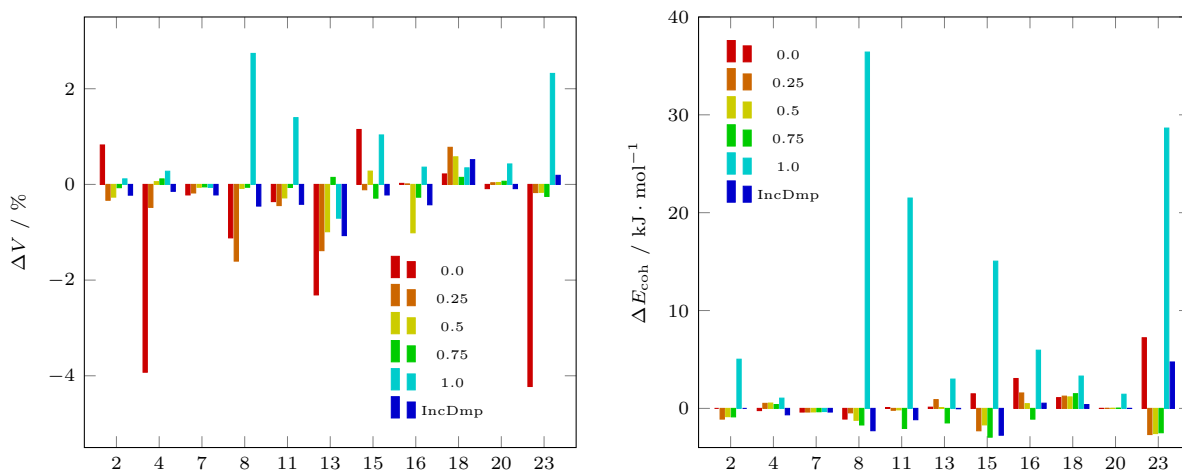


Figure 4.14: Cohesive energy differences ΔE_{coh} and cell volume differences ΔV for different damping factors s_{damp} , always calculated with respect to analogous calculations with dynamic damping. Results from ac(1)-QM/MM optimisations at the PBE-D3/TZVP+Chelpg+CGFF level of theory, restricted to molecules of the S11 subset. Not fully converged results are included.

A similar trend is found for the cell volumes. RMSDs to the reference are at most 1.6% for $s_{\text{damp}} < 1.0$, but an RMSD of 4.4% is obtained with $s_{\text{damp}} = 1.0$. This is, however, dominated by a single large deviation of 14.0% for cyanamide. For the other 10 systems an RMSD of only 1.1% is obtained. With Chelpg charges slightly larger RMSDs up to 2.0% for $s_{\text{damp}} < 1.0$ are found, and a slightly smaller RMSD of 1.1% for $s_{\text{damp}} = 1.0$. In V no systematic deviation for any $s_{\text{damp}} \neq 1.0$ is observed, whereas the $s_{\text{damp}} = 1.0$ results are always larger than the reference. Therefore, it seems plausible to assume that no systematic error is found in the reference dynamic damping results.

Since the results for $s_{\text{damp}} \neq 1.0$ include several geometry optimisations which have not fully been converged, it is not clear if the respective runs were already somewhat close to convergence when they broke down. Nevertheless, from these values it becomes obvious that even a strongly damped charge update has a significant influence on the results compared to no update.

Taking only the cases for which well converged results had been obtained without damping (ammonia (#4), carbon dioxide (#7), pyrazole (#18), and triazine (#20) for NPA charges) into account the findings are very similar. The largest cell volume and energy deviations are 2.6% and $8.1 \text{ kJ} \cdot \text{mol}^{-1}$ for pyrazole in the case of $s_{\text{damp}} = 1.0$. The deviations for $s_{\text{damp}} \neq 1.0$ are always significantly smaller, especially the energies which deviate by less than $0.5 \text{ kJ} \cdot \text{mol}^{-1}$ for these four systems. However, volume differences to the reference are similarly large for pyrazole both with no charge update and without damping. This might

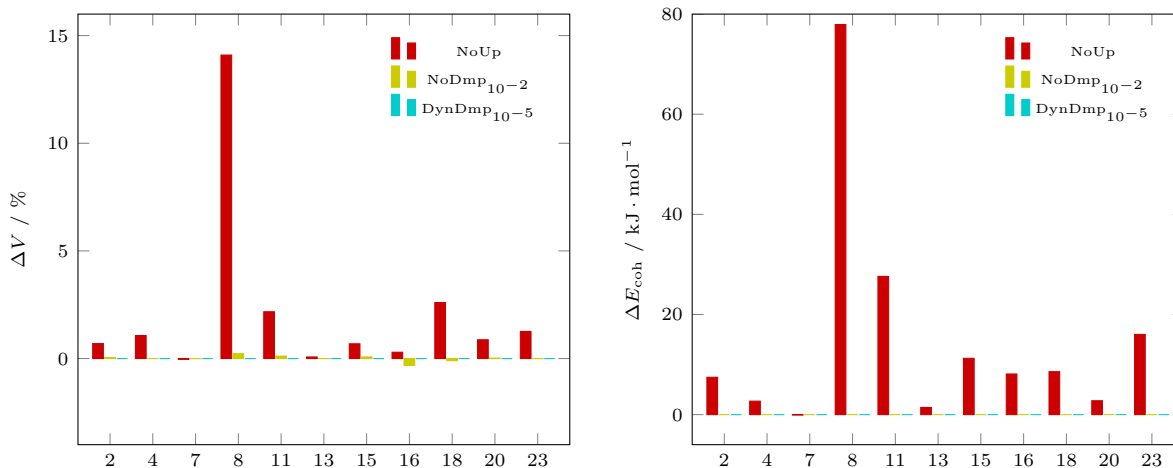


Figure 4.15: Cell volume deviation ΔV and cohesive energy deviation ΔE_{coh} for different charge update models with respect to the final charge update scheme. ac(1)-QM/MM calculations at the PBE-D3/TZVP+NPA+CGFF level of theory. No update of the charges (NoUp), an undamped charge update with a loose threshold of $q_{\text{thr}} = 10^{-2}$ (NoDmp_{10⁻²}), and a dynamic damping in conjunction with a rigid threshold of $q_{\text{thr}} = 10^{-5}$ e (DynDmp_{10⁻⁵}) are employed.

indicate that the dynamic damping in this case introduces some error.

The overall differences of the residual damping results compared to the dynamic damping results indicate a consistent behaviour of the latter. It is clear that the dynamic damping covers most of the impact which is introduced by a self-consistent determination of the electrostatic potential. Since cohesive energies are expected to be at most a few $\text{kJ} \cdot \text{mol}^{-1}$ too small, it is believed that the dynamic damping provides a much improved description compared to the neglect of a charge update, while enabling a stable geometry optimisation. Still, the convergence behaviour needs some improvement to be reliably usable for the whole X23 set.

4.6.4 Charge Damping with Different Thresholds

In order to make the update procedure more efficient (small number of QSCF cycles), but still robust (avoid convergence issues), the dynamic damping was employed in conjunction with a loose threshold of $q_{\text{thr}} = 10^{-2}$ e for the charge convergence. The threshold q_{thr} used for each calculation is denoted as a subscript to the damping method. For example, DynDmp_{10⁻²} refers to dynamic damping with a threshold q_{thr} of 10^{-2} e.

The results (Tables 4.24 and 4.25) now show a smooth convergence of all geometry optimisations. In contrast to the approaches modifying only the threshold without damping, or the damping with a tight threshold, no convergence problems are observed for the DynDmp_{10⁻²} update. Moreover, the number of optimisation steps until convergence is comparable to the calculations without any charge update. Only in the case of formamide with Chelpg charges a significant increase from 15 to 34 steps until convergence is observed. Nevertheless, in 10 of 11 cases with NPA and 9 of 11 cases with Chelpg the number of steps increases by less than 3. This is deemed an acceptable performance. The DynDmp_{10⁻²} update scheme reaches essentially the same performance (20.9 steps on average) as the calculations without update (20.8 steps on average) with NPA. Only slightly slower convergence (average of 20.4 steps with the update, average of 18.5 steps without) is found with Chelpg.

The differences of the DynDmp_{10⁻²} update (Figures 4.15 and 4.16) to those results using either only a loose threshold without damping or a dynamic damping with a tight threshold are marginal. Hence, just as for all previous damping results large differences to the NoUp case are obtained. Albeit the latter are considered to be less meaningful, the performance might still benefit from error cancellation.

Indeed, the comparison of cell volumes V and cohesive energies E_{coh} to experimental values (Table 4.26) shows a general trend between NoUp and the DynDmp_{10⁻²} update. Without any charge update volumes are

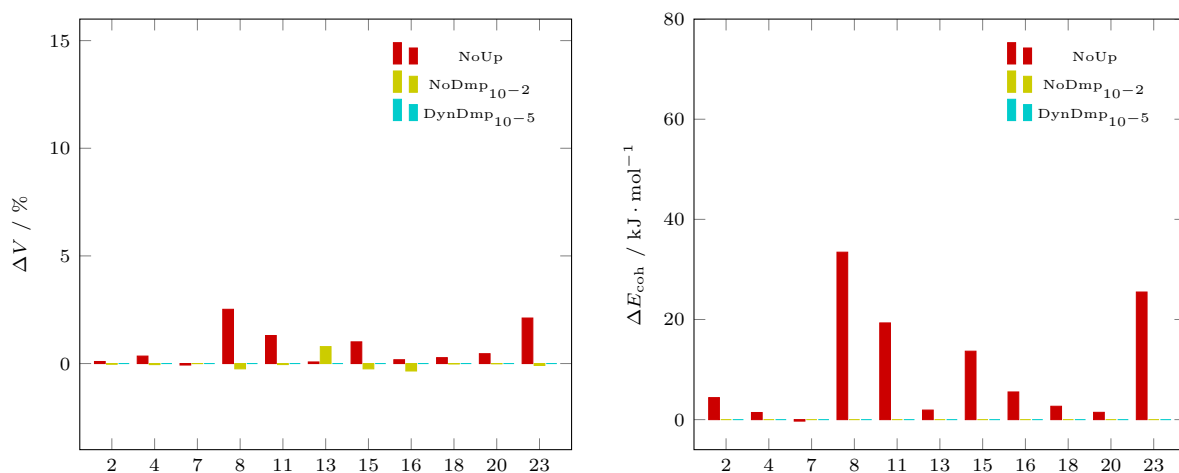


Figure 4.16: Cell volume deviation ΔV and cohesive energy deviation ΔE_{coh} for different charge update models with respect to the final charge update scheme. ac(1)-QM/MM calculations at the PBE-D3/TZVP+Chelpg+CGFF level of theory. No update of the charges (NoUp), an undamped charge update with a loose threshold of $q_{\text{thr}} = 10^{-2}$ (NoDmp_{10⁻²}), and a dynamic damping in conjunction with a rigid threshold of $q_{\text{thr}} = 10^{-5}$ e (DynDmp_{10⁻⁵}) are employed.

generally larger (mostly by 0 to 2%) and cohesive energies are less negative (i.e. weaker binding) than those obtained with the DynDmp_{10⁻²} update. Since the latter, compared to the experimental results, generally yields too small V and too negative E_{coh} , there seems to be an effective error compensation in the NoUp case. However, the possible improvement is in general small with respect to the absolute deviation, as can be seen from Figure 4.17. Moreover, three systems are not in line with this finding. In the cases of pyrazole and triazine, the optimised volumes are larger than the experimental values, thus, the results without update deviate more (4.8 and 9.0%) than the DynDmp_{10⁻²} ones (3.8 and 8.0%). The cyanamide results show much larger differences, namely a volume deviation of 14.1% and a cohesive energy deviation of $77.9 \text{ kJ} \cdot \text{mol}^{-1}$ in the NoUp case. While the DynDmp_{10⁻²} results are in line with the general trend of too small volumes (-4.5% with respect to experiment) and too negative cohesive energies ($-29.5 \text{ kJ} \cdot \text{mol}^{-1}$), the NoUp values show unreasonable deviations of 9.0% in V and $48.4 \text{ kJ} \cdot \text{mol}^{-1}$ in E_{coh} . Therefore, it is likely that the neglect of any charge update often counterbalances the overestimation of electrostatic interactions, but in some cases strongly overshoots. This causes an overall less reliable behaviour than the DynDmp_{10⁻²} update. Based on these results, the dynamic damping with a threshold of 10^{-2} e is used as the default throughout.

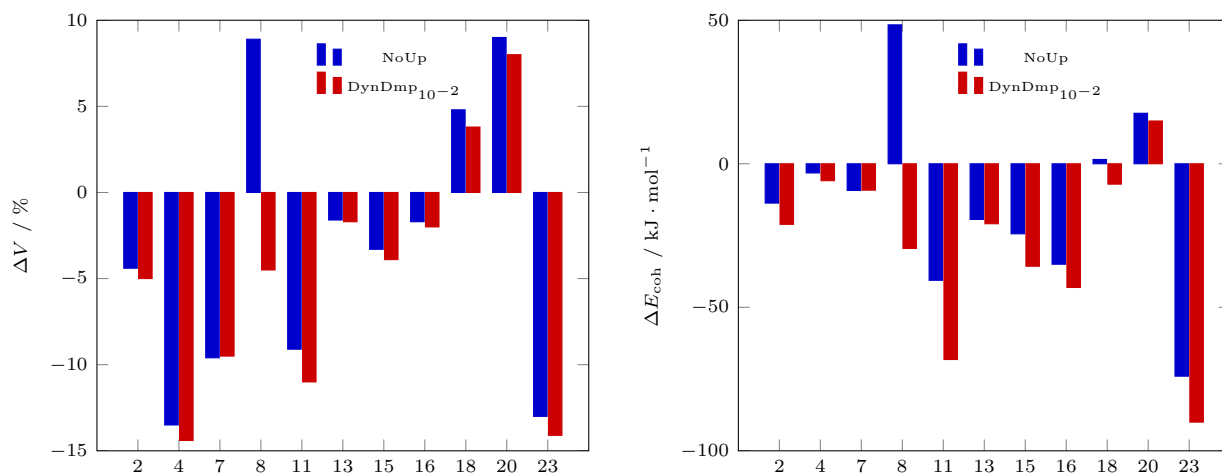


Figure 4.17: Cell volume deviation ΔV and cohesive energy deviation ΔE_{coh} from experimental values for ac(1)-QM/MM optimised structures at the PBE-D3/TZVP+NPA+CGFF level of theory, using either no charge update (NoUp) or the final update scheme of a dynamic damping with a loose convergence threshold (DynDmp_{10⁻²}).

Table 4.24: Convergence behaviour of the ac(1)-QM/MM geometry optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory with various charge update modifications. The number of optimisation steps until convergence is reached (or the maximum number of steps is reached) is given for calculations using no charge update (NoUp), an undamped charge update with a rigid threshold of $q_{\text{thr}} = 10^{-2}$ e (NoDmp_{10⁻²}), and with a dynamic damping in conjunction with a rigid threshold of $q_{\text{thr}} = 10^{-5}$ e (DynDmp_{10⁻⁵}) or a loose threshold of $q_{\text{thr}} = 10^{-2}$ e (DynDmp_{10⁻²})

	NoUp	NoDmp _{10⁻²}	DynDmp _{10⁻⁵}	DynDmp _{10⁻²}
acetic acid	17	19	18	18
ammonia	7	7	7	7
carbon dioxide	8	8	8	8
cyanamide	44	32	32	30
formamide	17	16	16	17
imidazole	32	29	30	29
oxalic acid (α)	26	27	26	28
oxalic acid (β)	27	(59)	(200)	37
pyrazole	26	27	27	27
triazine	8	8	8	8
urea	17	21	17	21
Mean(Tot)	20.8	23.0	35.4	20.9
Mean(Conv)	20.8	19.4	18.9	20.9

Table 4.25: Convergence behaviour of the ac(1)-QM/MM geometry optimisations at the PBE-D3/TZVP+Chelpg+CGFF level of theory with various charge update modifications. The number of optimisation steps until convergence is reached (or the maximum number of steps is reached) is given for calculations using no charge update (NoUp), an undamped charge update with a rigid threshold of $q_{\text{thr}} = 10^{-2}$ e (NoDmp_{10⁻²}), and with a dynamic damping in conjunction with a rigid threshold of $q_{\text{thr}} = 10^{-5}$ e (DynDmp_{10⁻⁵}) or a loose threshold of $q_{\text{thr}} = 10^{-2}$ e (DynDmp_{10⁻²})

	NoUp	NoDmp _{10⁻²}	DynDmp _{10⁻⁵}	DynDmp _{10⁻²}
acetic acid	20	28	24	25
ammonia	16	16	14	15
carbon dioxide	8	9	9	9
cyanamide	26	(65)	30	28
formamide	15	34	36	34
imidazole	14	31	34	14
oxalic acid (α)	25	23	26	23
oxalic acid (β)	36	(200)	39	33
pyrazole	18	18	20	18
triazine	8	9	8	8
urea	17	20	20	17
Mean(Tot)	18.5	41.2	23.6	20.4
Mean(Conv)	18.5	20.9	23.6	20.4

4.6.5 Background Charge Overfitting

Background charges are fitted to reproduce the electrostatic potential and electric field components at the atoms of the central molecule. For an N_a atomic molecule this corresponds to $4N_a$ values. If this number is smaller than the number N_{bq} of background charges, the resulting background charges suffer from overfitting. The influence could heavily be observed on the recalculation of an optimised geometry after a geometry optimisation. Final atomic coordinates and charges were read from disk to recalculate the background charges and the respective energies and gradients. The background charges almost sum up to 0 for both calculations and the sums are, up to two decimal places, identical. The RMS values, however, deviate significantly. As the values for background charges are not restricted in the fitting process, in some cases very large values can be obtained. Among all X23 systems, the RMS difference between the two sets of background charges is often (in 14 of 23 cases) larger than the RMS charge value.

Geometry optimisations and single-point calculations of the final structure were repeated several times on different machines to rule out random errors or inconsistencies among different hardware devices, and the results were observed to be exactly equal. These differences solely arise from the finite precision used to write the coordinates to disk after the geometry optimisation. In memory, coordinates and charges are stored as double precision numbers, which corresponds to approximately 16 decimal places in precision,

Table 4.26: Cell volumes V and cohesive energies E_{coh} from experiment (Exp.) and ac(1)-QM/MM geometry optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory, using no charge update (NoUp) or the combined update scheme (dynamic damping with a loose threshold, DynDmp₁₀₋₂). Volumes in \AA^3 , energies in $\text{kJ} \cdot \text{mol}^{-1}$.

	V			E_{coh}		
	Exp.	NoUp	DynDmp ₁₀₋₂	Exp.	NoUp	DynDmp ₁₀₋₂
acetic acid	297.3	284.2	282.3	-72.9	-86.6	-94.0
ammonia	135.1	116.8	115.6	-37.3	-40.5	-43.2
carbon dioxide	177.9	160.9	160.9	-27.2	-36.5	-36.4
cyanamide	415.6	452.9	396.9	-80.0	-31.6	-109.5
formamide	224.1	203.7	199.4	-79.1	-119.7	-147.3
imidazole	348.8	343.1	342.8	-87.1	-106.5	-108.0
oxalic acid (α)	312.6	302.4	300.3	-96.3	-120.7	-132.0
oxalic acid (β)	156.9	154.2	153.7	-96.3	-131.3	-139.4
pyrazole	698.3	731.8	724.5	-77.9	-76.4	-85.0
triazine	586.8	639.4	633.9	-62.0	-44.4	-47.1
urea	145.1	126.1	124.6	-102.6	-176.6	-192.6

but the values on disk only have 14 (coordinates) or 10 (charges) decimal places. Consequently, the values read from disk after the optimisation are not the same as those in memory at the end of the optimisation. Hence, the background charges fitted with these information are not the same. Since these tiny deviations in coordinates and charges induce huge changes in the background charges, the respective impact on energies and gradients has been monitored.

The comparison of root mean square (RMS) gradients from the final calculation of the optimisation run and from a subsequent single-point calculation with newly fitted background charges shows that the gradients are the same up to at least three decimal places (Table 4.27). All RMS gradients are in the order of 10^{-1} to $10^{-3} E_h \cdot a_0^{-1}$, while the root mean square deviations (RMSD) between both calculations are mostly of the order of less than $10^{-6} E_h \cdot a_0^{-1}$. The only exceptions are hexamine and urea ($3.0 \cdot 10^{-6}$ and $2.5 \cdot 10^{-6} E_h \cdot a_0^{-1}$, respectively). Since these values are still two orders of magnitude smaller than the RMS threshold of $9 \cdot 10^{-4} E_h \cdot a_0^{-1}$ used to determine the convergence of a geometry optimisation, the effect is considered negligible. For the energies of all 23 systems an RMSD of $5.3 \cdot 10^{-6} E_h$ is obtained. It is slightly above the convergence threshold of $3 \cdot 10^{-6} E_h$ in geometry optimisations. This is dominated by a single deviation of $25.2 \cdot 10^{-6} E_h$ in the case of hexamine, whereas the residual 22 molecules provide an RMSD of only $5.8 \cdot 10^{-8} E_h$. So, the maximum energy error from the change of background charges found within the X23 set is only $0.017 \text{ kJ} \cdot \text{mol}^{-1}$, which is absolutely negligible with respect to the possible accuracy of a QM/MM model.

However, it must be kept in mind that in rare cases the convergence of geometry optimisations might be hampered if the energy floats by a larger amount than the energy threshold. Albeit during any geometry optimisation the charges and coordinates are always saved in memory, the same effect may arise from very small geometry changes between steps. Even if the atomic charges do not change, the periodic potential used for the fitting procedure changes with any difference in atomic coordinates or cell parameters. Hence, the same problem may occur during a geometry optimisation and influence the convergence. In such cases one might need to adjust the number of background charges, or the cluster size in order to get rid of that effect. Nonetheless, from several hundreds of calculations there has been only one case in which this could be observed. Thus, the overfitting of background charges is deemed insignificant for the results obtained with the ac-QM/MM model and the background fitting procedure should be applicable as is.

Table 4.27: Root mean square (RMS) on the gradients taken from the final structure (Final) of an ac-QM/MM geometry optimisation at the PBE-D3/TZVP+NPA+CGFF level of theory and from a single-point calculation (Recalc.) of the final structure using the unchanged coordinates and atomic charges, but newly fitted background charges. The root mean square deviation (RMSD) and energy difference (ΔE_{coh}) are calculated between the respective gradients and energies of these two calculations. The total RMS or RMSD over the whole set is also given (Total). All gradients are given in $E_h \cdot a_0^{-1}$, all energies in E_h .

	RMS(Final)	RMS(Recalc.)	RMSD $\cdot 10^{-6}$	$\Delta E_{\text{coh}} \cdot 10^{-6}$
1,4-cyclohexanedione	0.014	0.014	0.026	0.003
acetic acid	0.150	0.150	0.059	0.073
adamantane	0.134	0.134	0.116	-0.026
ammonia	0.029	0.029	0.022	-0.015
anthracene	0.077	0.077	0.095	-0.002
benzene	0.004	0.004	0.024	0.007
carbon dioxide	0.018	0.018	0.052	0.002
cyanamide	0.113	0.113	0.140	-0.004
cytosine	0.096	0.096	0.012	-0.027
ethylcarbamate	0.096	0.096	0.051	0.000
formamide	0.119	0.119	1.260	-0.033
hexamine	0.009	0.009	5.189	25.162
imidazole	0.009	0.009	0.056	-0.001
naphthalene	0.094	0.094	0.005	0.004
oxalic acid (α)	0.042	0.042	0.559	-0.130
oxalic acid (β)	0.035	0.035	0.302	0.117
pyrazine	0.125	0.125	0.002	-0.002
pyrazole	0.124	0.124	0.022	-0.003
succinic acid	0.016	0.016	0.047	0.017
triazine	0.109	0.109	0.002	0.001
trioxane	0.024	0.024	0.003	0.006
uracil	0.151	0.151	0.135	-0.176
urea	0.012	0.012	4.352	0.065
Total	0.086	0.086	1.445	5.247

4.7 Two-body ac-QM/MM

While ac(1)-QM/MM can already provide reasonable geometries for most benchmark systems, it is clear that strong interactions are not described properly at a QM/MM level. Albeit a couple of methods would be available to improve the QM/MM interactions - point charges to be replaced with distributed multipoles, LJ parameters fitted self-consistently from QM calculations, consideration of exchange-repulsion potentials for the QM Hamiltonian - these could severely hamper the applicability. Since the required parameters are not routinely available, whereas point charges or force field parameters are easily accessible and exchangeable, an approach based on QM interactions seems favourable. Therefore, the inclusion of intermolecular interactions from a supramolecular approach was preferred. Unfortunately, a simple expansion of the QM region to e.g. the first shell of molecules can quickly become computationally too demanding due to the scaling of computational cost with system size even for GGA DFT. For that reason, an additive approach, calculating QM interactions for molecule pairs one after another, is the method of choice.

4.7.1 Selection of the Two-body Region

The QM molecule pair interactions are calculated for a limited set of molecules. These molecules form the two-body region R . For a balanced description two-body corrections need to be included for a sufficiently large, but preferably small number of surrounding molecules. Since the cluster building routine uses a cutoff distance from the origin point, it seems reasonable to apply a similar criterion for the selection of a two-body region. For each molecule m the cartesian mean of its atom coordinates is determined as

$$\vec{r}_{\text{mean}}(m) = \sum_{A \in m} \vec{r}(A). \quad (4.58)$$

If it lies within a cutoff distance r_R from the cluster origin, the molecule is added to the two-body region R .

The cutoff distance is determined as follows. From the cell volume V and the number of molecules per cell N_{cell} the volume per molecule V_m is calculated. For a cubic molecule this V_m would correspond to an edge length of $\sqrt[3]{V_m}$. Thus, a cutoff selection criterion

$$r_{\text{mean}} \leq 1.25 \cdot S_{\text{sh}} \cdot \sqrt[3]{V_m} \quad (4.59)$$

is chosen with the arbitrary factor 1.25 to account for the non-cubic shape of the molecule and the factor $S_{\text{sh}} \geq 1$ to account for the size of the region. This is referred to as the centre distance approach (CentDist).

A second approach does not use mean coordinates, but explicit atom-atom distances. For each molecule m with atoms B the minimum molecular distance

$$r_{\text{min}}(m) = \min(|\vec{r}(m, B) - \vec{r}(0, A)|) \quad (4.60)$$

is defined as the minimum of all pairwise atom distances $|\vec{r}(m, B) - \vec{r}(0, A)|$ to the central molecule 0 with atoms A .

Molecules are then included if the condition

$$r_{\text{min}}(m) \leq r_R \quad (4.61)$$

with the cutoff parameter r_R is fulfilled. A value of $r_R = 6 \text{ a}_0$ was found to be a reasonable choice in many cases, but larger values are required for strongly interacting systems. Hence, $r_R = 6 \text{ a}_0$ is suggested (and implemented) as a default, but it seems much more reasonable to use system specific cutoffs. For the X23 set a series of r_R was determined from preliminary tests. A stepwise increase of r_R until reasonable convergence of the ac(2)-QM/MM energy to $2 \text{ kJ} \cdot \text{mol}^{-1}$ is used in order to keep the two-body region as small as possible. The list of computed r_R parameters is included in Table 4.29. These parameters were used throughout all

ac(2)-QM/MM calculations. This procedure guarantees converged two-body regions R for molecules that strongly deviate from a spherical shape. It is referred to as the atom distance approach (AtomDist).

4.7.2 General Formalism ac(2)-QM/MM

At the ac(1)-QM/MM level, both the internal energy of the central molecule and its interactions with the environment are taken from a single embedded calculation as

$$E_{\text{ac(1)-QM/MM}} = E_{\text{emb}}^{1\text{B}}(0). \quad (4.62)$$

In order to introduce two-body corrections on top of the ac(1)-QM/MM result, it is required to remove the QM/MM based interaction with a limited number of molecules, and add the respective QM interactions. It is based on a standard supramolecular approach for pairwise interaction energies

$$\Delta E(mn) = E(mn) - E(m) - E(n). \quad (4.63)$$

This leads to an expression for the two-body energy of the form

$$E_{\text{ac(2)-QM/MM}} = E_{\text{ac(1)-QM/MM}} - \sum_{m \in R} E_{\text{emb}}^{1\text{B}}(0) - E_{\text{emb}\setminus m}^{1\text{B}}(0) + \sum_{m \in R} E_{\text{emb}}^{2\text{B}}(0m) - E_{\text{emb}\setminus m}^{2\text{B}}(0) - E_{\text{emb}\setminus 0}^{2\text{B}}(m), \quad (4.64)$$

where $E_{\text{emb}}^{1\text{B}}(0)$ and $E_{\text{emb}\setminus m}^{1\text{B}}(0)$ are from embedded calculations with and without molecule m present in the embedding. $E_{\text{emb}}^{2\text{B}}(0m)$ is the energy of combined molecules 0 and m in their full embedding potential, $E_{\text{emb}\setminus m}^{2\text{B}}(0)$ is the energy of molecule 0 embedded without m , and $E_{\text{emb}\setminus 0}^{2\text{B}}(m)$ is the energy of molecule m embedded without molecule 0. The first sum removes the QM/MM interactions and the second sum adds the QM interactions. In this form the scheme requires a total of four QM calculations for each molecule within R , while the $E_{\text{emb}}^{1\text{B}}(0)$ term is the result from an ac(1)-QM/MM calculation, see Equation 4.62. Nevertheless, the demand can be reduced if two-body corrections are calculated at the same level of theory as the one-body results. In that case the former terms $E_{\text{emb}\setminus m}^{1\text{B}}(0)$ and $E_{\text{emb}\setminus m}^{2\text{B}}(0)$ become equal and cancel out, and the residual terms do not need to be distinguished from the theory level. Therefore, Equation 4.64 reduces to

$$E_{\text{ac(2)-QM/MM}} = E_{\text{ac(1)-QM/MM}} + \sum_{m \in R} E_{\text{emb}}(0m) - E_{\text{emb}}(0) - E_{\text{emb}\setminus 0}(m), \quad (4.65)$$

which requires only two QM calculations for each molecule within R . All calculations including two-body corrections in this manner are denoted as ac(2)-QM/MM, or shortly ac(2) within tables and figures.

4.7.3 Variation ac(2-F)-QM/MM

One approach, referred to as ac(2-F)-QM/MM, is based on the assumption that over-polarisation might be a significant issue in the ac-QM/MM model. According to Equation 4.64 the energy $E_{\text{emb}\setminus m}^{1\text{B}}(0)$ of monomer 0 is calculated without the charges of monomer m , using a converged wave-function. But instead of the latter, the wave-function may be calculated with the charges of m present, and the energy can then be evaluated with this wave-function and the charges of m absent. So, the energy expression is slightly changed to

$$E_{\text{ac(2-F)-QM/MM}} = E_{\text{ac(1)-QM/MM}} - \sum_{m \in R} E_{\text{emb}}^{1\text{B}}(0) - E_{\text{emb}\setminus m}^{1\text{B}*}(0) + \sum_{m \in R} E_{\text{emb}}^{2\text{B}}(0m) - E_{\text{emb}\setminus m}^{2\text{B}}(0) - E_{\text{emb}\setminus 0}^{2\text{B}}(m) \quad (4.66)$$

where $E_{\text{emb}\setminus m}^{1\text{B}*}(0)$ is now the energy of molecule 0 calculated with no charges at molecule m , but using the unchanged wave-function of a calculation with charges of m present. As for the unchanged scheme, one- and two-body theory levels can be chosen to be the same. In that case, and in combination with Equation 4.65,

one obtains

$$E_{\text{ac}(2\text{-F})\text{-QM/MM}} = E_{\text{ac}(2)\text{-QM/MM}} + \sum_{m \in R} \left(E_{\text{emb}\setminus m}^*(0) - E_{\text{emb}\setminus m}(0) \right). \quad (4.67)$$

The additional term represents the reorganisation energy induced by the charges of m .

However, a major drawback of this construction is the limitation to energies. Since the calculation of $E_{\text{emb}\setminus m}^*(0)$ uses a wave-function which had been converged for a different set of charges, it is not converged for the energy calculation. Therefore, gradients can not be obtained from the ac(2-F)-QM/MM scheme. This variant may only be applied as an a posteriori energy correction, or it must rely on numerical gradients.

4.7.4 Variation ac(2-I)-QM/MM

Another modification, referred to as ac(2-I)-QM/MM, modifies the calculation of $E_{\text{emb}\setminus 0}(m)$. Firstly, it is assumed that instead of Equation 4.65 it should also be possible to use

$$E_{\text{ac}(2\text{-I})\text{-QM/MM}} = E_{\text{ac}(1)\text{-QM/MM}} + \sum_{m \in R} E_{\text{emb}}(0m) - E_{\text{emb}\setminus m}(0) - E_{\text{emb}}(m). \quad (4.68)$$

In this case the QM/MM electrostatic interactions between molecules 0 and m are calculated for the QM molecule m with charges at molecule 0. Instead of a single calculation without charges and LJ parameters at molecule m for $E_{\text{emb}\setminus m}(0)$, two different calculations are combined. One calculation uses the same charges on molecule m as on all other cluster molecules (which is the same as $E_{\text{ac}(1)\text{-QM/MM}}$, so, no extra calculation is required), the second uses charges of the same magnitude, but with the opposite sign. The respective energies for these are $E_{\text{emb}}(m)$ and $E_{\text{emb}[m]}(m)$. The interactions with molecule 0 should to great extent cancel out for

$$E_{\text{emb}\setminus m}^{(\text{approx})}(0) = \frac{1}{2} \left(E_{\text{emb}}(0) + E_{\text{emb}[m]}(0) \right). \quad (4.69)$$

If the approximate energy $E_{\text{emb}\setminus m}^{(\text{approx})}(0)$ benefits from error compensation, one might obtain improved energies with the ac(2-I)-QM/MM approach. This variant allows for the straightforward use of analytical gradients.

4.7.5 Performance of the Two-body Region Selection

For all X23 molecules, ac(2)-QM/MM geometry optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory were carried out using the centre distance or atom distance selection criteria for the two-body region. For the selection according to molecule centre distances (Equation 4.59) scaling factors S_{sh} of 1 and 2 were used. If the selection according to the explicit atom-atom distance (Equation 4.60) was used, either a uniform cutoff distance of 6 a_0 , or the predetermined cutoff distance according to Table 4.29 was applied.

The results (Figure 4.18) with uniform small cutoffs ($S_{\text{sh}} = 1$ and $r_R = 6 a_0$) for both selection schemes show an overall trend in both V and E_{coh} . For several systems the selections according to $S_{\text{sh}} = 1$ and $r_R = 6 a_0$ yield similar results. Differences of less than 1% in the volume and less than 4 $\text{kJ} \cdot \text{mol}^{-1}$ in the cohesive energy are observed for almost one half of the systems. In some cases, however, significantly larger differences between the results are observed. For instance, succinic acid and uracil show differences of 41.2 and 32.1 $\text{kJ} \cdot \text{mol}^{-1}$ in the cohesive energy, and cyanamide and pyrazine deviations of 3.2 and 5.2% in the cell volume. In five cases (1,4-cyclohexanedione, anthracene, formamide, naphthalene, and pyrazole) the centre based selection leads to cell volumes closer to experiment than the atom distance based, but the difference between both selection schemes are small (0.4, 0.1, 0.3, 0.1 and 0.7% in these cases). A similar behaviour is observed for the cohesive energies, which are for some systems closer to experiment if the centre based approach is used. However, the CentDist based energies are regularly by less than 3 $\text{kJ} \cdot \text{mol}^{-1}$ better than those from the AtomDist approach, with the only exception of trioxane (5.9 $\text{kJ} \cdot \text{mol}^{-1}$). So, overall there are no systems for which the centre distance approach performs notably better than the atom distance approach, but several systems for which the opposite is observed. This strongly indicates that the atom distance approach is more reasonable and should be used for subsequent calculations.

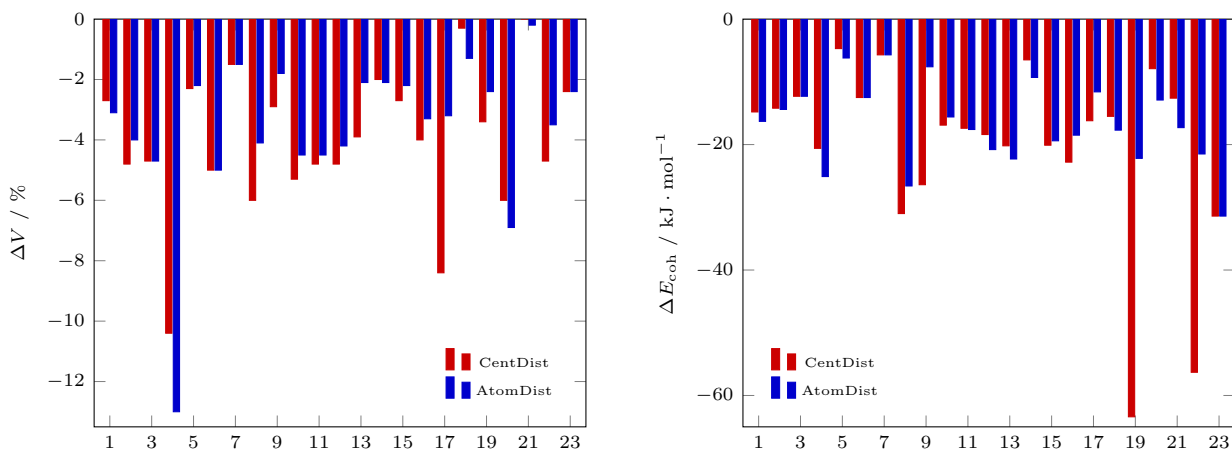


Figure 4.18: Deviation of optimised cell volumes V and cohesive energies E_{coh} from experimental values for ac(2)-QM/MM calculations at the PBE-D3/TZVP+NPA+CGFF level of theory. Two-body regions were selected according to either the centre distance approach with $S_{\text{sh}} = 1$ (CentDist) or the pairwise atom distance approach with $r_R = 6 a_0$ (AtomDist).

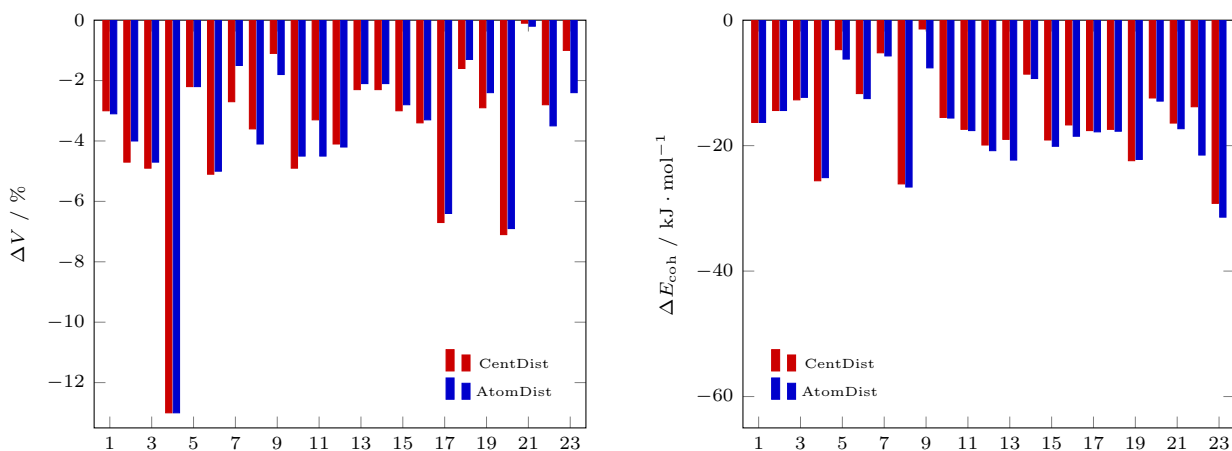


Figure 4.19: Deviation of optimised cell volumes V and cohesive energies E_{coh} from experimental values for ac(2)-QM/MM calculations at the PBE-D3/TZVP+NPA+CGFF level of theory. Two-body regions were selected according to either the centre distance approach with $S_{\text{sh}} = 2$ (CentDist) or the pairwise atom distance approach with $r_R = r_R^f$ (AtomDist).

The differences are greatly reduced if S_{sh} is increased to 2, and the pre-determined r_R values for all systems are used. The results (Figure 4.19) of these sets are very similar in volumes and cohesive energies. Differences in the cell volumes between the sets are mostly below 1%, although differences up to 1.4% (urea) are found. However, it can not be observed that one of the selection criteria yields systematically smaller differences with respect to experimental results. The same behaviour is found for the cohesive energies. Between both sets, most differences are well below $1 \text{ kJ} \cdot \text{mol}^{-1}$. These are most likely well converged with both selection criteria. Nevertheless, in few cases significant differences up to $7.7 \text{ kJ} \cdot \text{mol}^{-1}$ (uracil) are observed. In these cases (cytosine, uracil), it is not clear whether the results are well converged with the size of the two-body region. These cases are investigated in more detail (see Section 4.7.6). While both approaches perform very similar with these parameters, the efficiency is very different. With pre-determined r_R parameters at most 24 molecules are included in the two-body region, and on average only 14.0 molecules. The centre distance approach with $S_{\text{sh}} = 2$ includes 54 to 74 molecules, 64.5 on average. Therefore, the number of required QM calculations for the latter is roughly 4.5 times larger. Although one might optimise the S_{sh} parameter, the reasonable performance of the atom distance approach with $r_R = 6 a_0$ in many cases encourages the use of this selection approach.

Another indicator for the superiority of the atom distance approach is the shape of the two-body region which is obtained in some cases, e.g. cyanamide. It can be seen from Figure 4.20 that partially different molecules are included in the two-body region. The centre distance approach includes fewer molecules (7) compared to the atom distance approach (12), but is not simply a subset of the latter one. The overall shape

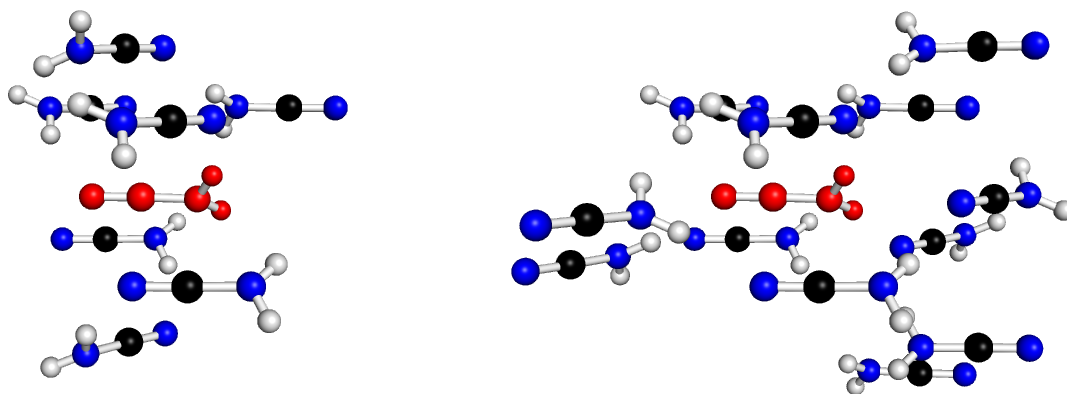


Figure 4.20: Two-body regions around the central molecule, coloured in red. The molecules were selected according to the centre distance approach with $S_{\text{sh}} = 1$ (left) or the pairwise minimum atom distance approach with $r_R = 6 a_0$ (right).

is distinctly different. While the centre distance scheme forms a ring of molecules around the cyanamide axis, even including two more molecules in that region, the atom distance approach creates a somewhat uniform surrounding, especially covering the “edges” of the central molecule. The latter is believed to be a more reasonable choice for the molecules to calculate corrections for, which is in line with the previous findings.

Albeit it is not assured that either of the approaches has reached convergence in the energy correction, it is still deemed reasonable to choose the atom distance approach as the default for the selection of the two-body region.

4.7.6 Convergence Behaviour with the Two-body Region Size

In order to obtain a set of reasonable r_R values for the selection according to Equation 4.60, the parameter was varied from 6 to at least 10 a_0 in steps of 1 a_0 . In order to save computational time, full geometry optimisations at the ac(2)-QM/MM PBE-D3/TZVP+NPA+CGFF level of theory were only performed for a uniform value of $r_R = 6 a_0$. Energies for all residual r_R values were calculated on the structure optimised with $r_R = 6 a_0$.

The resulting cohesive energies (Table 4.28) mostly show small changes of few $\text{kJ} \cdot \text{mol}^{-1}$ with the variation of r_R . Nevertheless, in several cases the cutoff increase from 6 to 7 or 8 a_0 leads to somewhat large differences in E_{coh} , e.g. for ammonia, formamide, or pyrazine (-3.4 , 3.2 , and $4.8 \text{ kJ} \cdot \text{mol}^{-1}$, respectively). Most systems converge reasonably to a limit, as can be seen from the examples in Figure 4.21. Although sometimes small oscillations in the order of $1 \text{ kJ} \cdot \text{mol}^{-1}$ can be observed, for instance in the case of formamide (values of -99.6 , -96.4 , -97.1 , -96.4 , and $-97.1 \text{ kJ} \cdot \text{mol}^{-1}$). Nevertheless, these results can be deemed well converged to below $1 \text{ kJ} \cdot \text{mol}^{-1}$, leaving only two difficult cases, namely cytosine and uracil. The former displays a surprising jump of $3.1 \text{ kJ} \cdot \text{mol}^{-1}$ in the cohesive energy when r_R is increased from 9 to 10 a_0 , and results for even larger selection cutoff distances of 11 and 12 a_0 still cause rather large changes of 0.6 and $2.0 \text{ kJ} \cdot \text{mol}^{-1}$. It seems that a somewhat converged energy of about $-170 \text{ kJ} \cdot \text{mol}^{-1}$ is reached at $r_R = 14 a_0$, for large two-body regions the energies more or less oscillate (-170.9 , 168.9 , 170.4 , and $169.9 \text{ kJ} \cdot \text{mol}^{-1}$ for r_R of 11, 12, 14, and 16 a_0). Similarly, for uracil quite constant changes of the cohesive energy occur when r_R is altered from 6 to 10 a_0 , and only at 14 a_0 it seems that convergence to about $-150 \text{ kJ} \cdot \text{mol}^{-1}$ has been reached. However, in both cases the deviation between calculations with a small r_R of 7 a_0 and a much larger value of 16 a_0 are only about $6 \text{ kJ} \cdot \text{mol}^{-1}$. The required number of dimer calculations increases from 15 to 47 (cytosine) or from 15 to 69 (uracil). While these would still have been feasible with the TZVP basis, the expected slowdown of further benchmarking calculations with larger basis sets was deemed unacceptable. Therefore, it was decided to restrain r_R to 7 a_0 for all further ac(2)-QM/MM calculations on these two systems, and to keep the error in mind.

Table 4.28: Cohesive energies from single-point ac(2)-QM/MM calculations at the PBE-D3/TZVP+NPA+CGFF level of theory. The selection of a two-body region was performed according to the atom distance criterion with different values for the selection cutoff r_R . This was always varied from 6 up to 10 a_0 , in difficult cases it was further varied up to 16 a_0 . All energies in $\text{kJ} \cdot \text{mol}^{-1}$.

	6 a_0	7 a_0	8 a_0	9 a_0	10 a_0	11 a_0	12 a_0	14 a_0	16 a_0
1,4-cyclohexanedione	-105.1	-104.7	-106.0	-106.0	-106.0	-	-	-	-
acetic acid	-87.2	-87.8	-88.1	-87.6	-87.6	-	-	-	-
adamantane	-81.8	-81.8	-82.2	-82.4	-82.4	-	-	-	-
ammonia	-57.9	-57.9	-61.3	-62.3	-62.6	-	-	-	-
anthracene	-118.9	-119.6	-119.6	-119.6	-119.0	-	-	-	-
benzene	-64.5	-64.5	-64.5	-64.4	-64.0	-	-	-	-
carbon dioxide	-32.9	-32.9	-32.9	-32.9	-32.8	-	-	-	-
cyanamide	-108.8	-105.3	-105.0	-105.2	-105.1	-	-	-	-
cytosine	-174.1	-176.2	-176.9	-174.6	-171.5	-170.9	-168.9	-170.4	-169.9
ethylcarbamate	-101.8	-102.9	-102.7	-102.3	-102.0	-	-	-	-
formamide	-99.6	-96.4	-97.1	-96.4	-97.1	-	-	-	-
hexamine	-107.0	-107.0	-107.0	-107.0	-107.0	-	-	-	-
imidazole	-109.4	-109.1	-109.2	-109.3	-107.3	-	-	-	-
naphthalene	-90.9	-91.6	-91.6	-91.7	-91.4	-	-	-	-
oxalic acid (α)	-116.4	-116.5	-117.3	-116.5	-116.5	-	-	-	-
oxalic acid (β)	-114.8	-115.3	-113.7	-113.7	-113.6	-	-	-	-
pyrazine	-73.2	-78.0	-78.0	-78.0	-78.5	-	-	-	-
pyrazole	-93.4	-94.4	-95.6	-95.6	-95.9	-	-	-	-
succinic acid	-152.4	-152.4	-152.4	-152.3	-152.2	-	-	-	-
triazine	-69.5	-73.7	-73.7	-73.7	-73.7	-	-	-	-
trioxane	-85.1	-85.1	-85.1	-83.7	-83.7	-	-	-	-
uracil	-159.8	-156.4	-154.7	-154.8	-154.0	-151.6	-151.4	-149.1	-150.4
urea	-134.0	-134.0	-134.2	-134.2	-132.5	-	-	-	-

The final parameters r_R^f were determined and are provided in Table 4.29. The values were selected by monitoring the changes in E_{coh} . If an increase in r_R led to $\Delta E_{\text{coh}} < 2 \text{ kJ} \cdot \text{mol}^{-1}$ (with a change in the number of molecules in the two-body region) the value was kept. In some calculations the optimisation and the energy were computed with different values. For example, R(7)@R(6) will denote an optimisation with the two-body region selected according to $r_R = 6 a_0$, but a final energy calculation with a selection according to $r_R = 7 a_0$.

Interestingly, the number of molecules in the converged two-body region is quite similar for most systems. Its size mainly ranges from 10 to 15 molecules, with a mean of 14.0, and only ammonia stands out with a shell of 30 molecules. Since ammonia provides several intermolecular hydrogen bonds and, thus, short intermolecular distances, it seems quite reasonable to require a larger number of neighbouring molecules to reach convergence. Surprisingly, no trend is observed for hydrogen bonded systems to require a larger number of molecules in the two-body region. The average number of 14.4 is only slightly higher than the total average. The selections based on the same distance criterion, but either applied to the $R(6)$ or $R(r_R^f)$ optimised structure, are in 3 out of 9 cases (ammonia, cytosine, and pyrazole) different. This indicates molecules to have moved into or out of the cutoff distance during the optimisation, which might in principle affect the results.

Nevertheless, the energies are not strongly changed upon the selection. The $R(r_R^f)$ @ $R(6)$ energies may be either lower or higher than the $R(6)$ @ $R(6)$ values, dependent on the direction the additional correction. As expected, for all 9 systems the $R(r_R^f)$ energies obtained from the $R(r_R^f)$ geometries are a little lower than those from the $R(6)$ geometries, but only by $0.7 \text{ kJ} \cdot \text{mol}^{-1}$ on average. This is the case for all systems, no matter if the selection is the same. Consequently, it is concluded that the final r_R^f parameter determination based on the $R(6)$ structures is reasonable.

It might be assumed that in general an optimisation with a small two-body region and final energy evaluation for a larger region might be useful. This would require the energy landscape to be similar for a small and a large selection. Then the main contribution stems from the interactions of the additional molecules in the larger region, but only a small part from the geometry difference of the optimised structures.

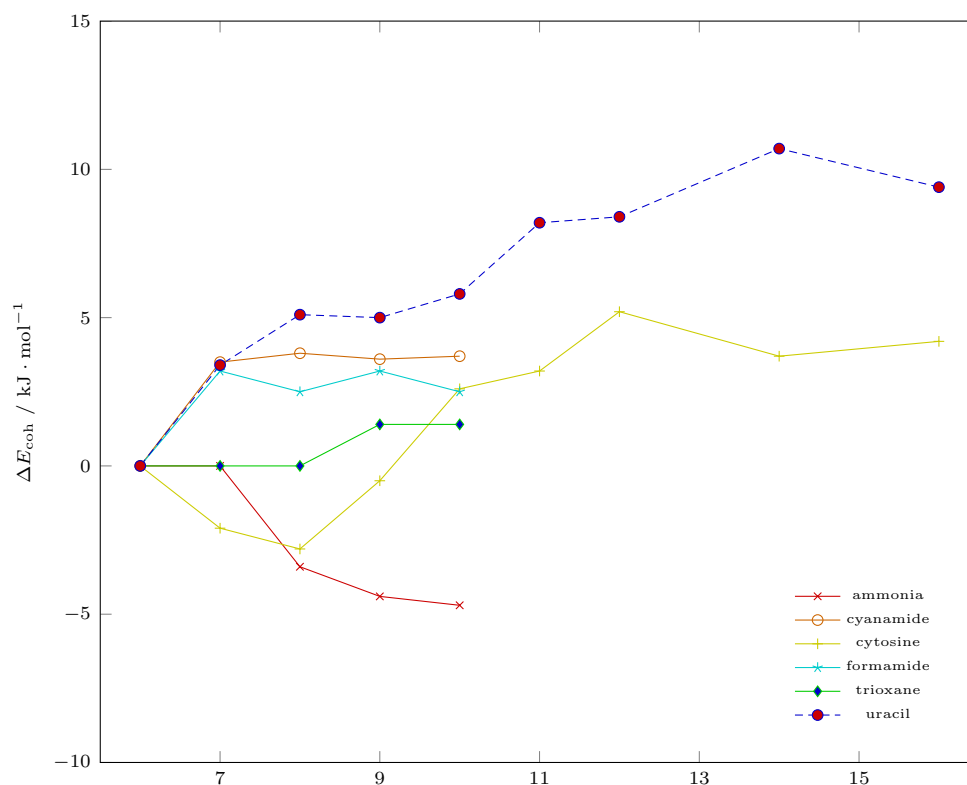


Figure 4.21: Change of the cohesive energy with the size of the two-body region, represented with the atom distance selection criterion r_R . Energies are given relative to the results obtained with $r_R = 6 a_0$.

Unless some strongly interacting molecules are left out of the two-body region, it is believed to be a reasonable approach. Especially for larger molecules and basis sets, when dimer calculations become very expensive, this will become useful. Although this seems to be working very well, based on the X23 results presented above, it was not yet used as the default procedure. For all subsequent results the optimisation and energy calculations were run using the same definition of the two-body region.

4.7.7 Applicability of a Dimer Approach

In this section, we look in further detail into the intermolecular interactions used for the ac(2)-QM/MM approach. Cyanamide was chosen as a test subject, because it had shown an erratic behaviour during some preliminary tests. Interactions between the central molecule (0) and the seven next neighbouring molecules (1-7, Figure 4.22) from the cluster were taken into account, as these are expected to show both very strong interactions and a large impact on the respective many-body corrections to the energy.

Gas Phase Interaction Energies

For each pair of the central molecule with one of its seven neighbours the DFT-SAPT interaction energy at the PBE/TZVP level of theory was calculated. DFT-SAPT is deemed a reasonable reference approach from the DFT side for interaction energies, especially since it is not biased from the intermolecular BSSE, which contributes to intermolecular interaction energies. On the other hand, the use of wave-function based techniques would allow for an improved description, thus, these might help to identify differences arising from the supramolecular description itself. Specifically the SCS-LMP2 variant should yield interaction energies close to DFT-SAPT. Different wave-funtion calculations were performed in order to assure that these are in line with the DFT-SAPT approach. Supramolecular interaction energies were calculated from PBE without and with different approaches for dispersion interactions, with different variants of the MP2 technique, and with the B2PLYP method. All calculations were carried out with the TZVP basis set.

Table 4.29: Final r_R^f parameters used for the X23 molecules, the respective number of molecules in the two-body region N_m , and the cohesive energy of calculations with $R(6)$ and $R(r_R^f)$. If $r_R^f \neq 6$ a₀, the cohesive energies calculated with r_R^f are given for both the $R(6)$ and $R(r_R^f)$ optimised structures. All r_R in a₀, all energy differences in kJ · mol⁻¹.

	r_R^f	N_m	$R(6)@R(6)$	$R(r_R^f)@R(6)$	$R(r_R^f)@R(r_R^f)$
1,4-cyclohexanedione	6	12	—	—	—
acetic acid	6	12	—	—	—
adamantane	6	12	—	—	—
ammonia	9	30	-57.9	-62.3	-62.3
anthracene	6	12	—	—	—
benzene	6	12	—	—	—
carbon dioxide	6	12	—	—	—
cyanamide	7	15	-108.8	-105.3	-106.5
cytosine	7	15	-174.1	-176.2	-177.6
ethylcarbamate	6	13	—	—	—
formamide	7	14	-99.6	-96.4	-96.8
hexamine	6	14	—	—	—
imidazole	6	13	—	—	—
naphthalene	6	12	—	—	—
oxalic acid (α)	6	12	—	—	—
oxalic acid (β)	6	10	—	—	—
pyrazine	7	14	-73.2	-78.0	-79.4
pyrazole	8	15	-93.4	-95.6	-95.6
succinic acid	6	14	—	—	—
triazine	7	14	-69.5	-73.7	-74.9
trioxane	9	20	-85.1	-83.7	-83.9
uracil	7	14	-159.8	-156.4	-157.2
urea	6	10	—	—	—
Mean	—	14.0	—	—	—

Table 4.30: Gas phase interaction energies between pairs (0 + i) of the central molecule (0) and the next molecules i (1 to 7) from the cyanamide crystal. DFT-SAPT calculations at the PBE/TZVP level of theory, supramolecular calculation (Supra) at different MP2 derived levels of theory. All values in kJ · mol⁻¹.

	DFT-SAPT			Supra	
	PBE/TZVP	MP2/TZVP	SCS-MP2/TZVP	SCS-LMP2/TZVP	B2PLYP/TZVP
0+1	11.3	7.0	8.8	10.2	11.8
0+2	11.3	7.0	8.8	10.2	11.8
0+3	-16.5	-19.1	-17.3	-15.7	-15.3
0+4	-16.5	-19.1	-17.3	-15.7	-15.3
0+5	-3.0	-4.6	-3.6	-2.8	-2.6
0+6	-26.4	-29.8	-27.6	-26.0	-27.7
0+7	-26.4	-29.8	-27.6	-26.0	-27.7
Sum	-66.3	-88.3	-75.9	-65.8	-64.9

The results for all molecule pairs (Table 4.30) show well-known trends among the theoretical methods employed. MP2 and SCS-MP2 both yield more stabilising energies than DFT-SAPT, but SCS-LMP2 cohesive energies are in excellent agreement. MP2 is expected to suffer from an overestimation of dispersion, which is reduced in the SCS-MP2 scheme. The intrinsic influence of the intermolecular BSSE should be compensated from the local approach in SCS-LMP2. Consequently, the sum of the interaction energies becomes less binding from MP2 (-88.3 kJ · mol⁻¹) via SCS-MP2 (-75.9 kJ · mol⁻¹) to SCS-LMP2 (-65.8 kJ · mol⁻¹). The latter is in excellent agreement with DFT-SAPT (-66.3 kJ · mol⁻¹), with largest deviations of 1.1 kJ · mol⁻¹ and an RMSD of 0.7 kJ · mol⁻¹. The perturbational approach B2PLYP behaves very similarly with a slightly larger RMSD of 1.4 kJ · mol⁻¹ with respect to DFT-SAPT. The results support our ansatz and confirm DFT-SAPT, SCS-LMP2, and B2PLYP as reliable levels of theory.

Using DFT-SAPT as a reference for the further supramolecular interactions energies (Table 4.31) the relative performance of PBE and PBE-D3 is quite unexpected. The supramolecular PBE-D3 differences to DFT-SAPT yield an RMSD of 4.9 kJ · mol⁻¹ with a maximum deviation of 6.0 kJ · mol⁻¹, but the PBE energy differences are only 1.2 kJ · mol⁻¹ for the RMSD and 1.8 kJ · mol⁻¹ for the maximum deviation.

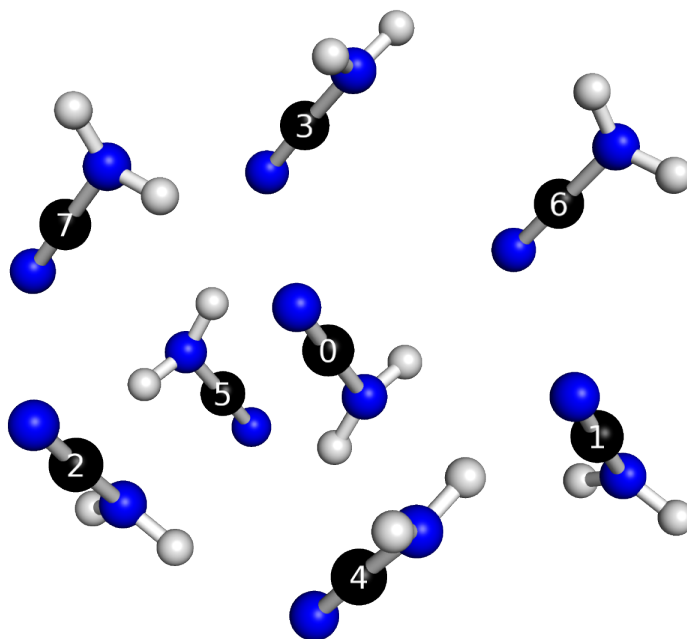


Figure 4.22: The central molecule (0) and its seven next neighbours in the cyanamide crystal, selected using the centre distance approach.

As DFT-SAPT is expected to cover dispersion contributions, it seems strange to find the better agreement without the D3 dispersion corrections. However, these findings are supported from the non-local DFT calculations at the PBE/TZVP level of theory. These provide interaction energies that are very similar to the PBE-D3 ones. A very similar RMSD of $4.5 \text{ kJ} \cdot \text{mol}^{-1}$ and maximum deviation of $5.9 \text{ kJ} \cdot \text{mol}^{-1}$ with respect to the reference DFT-SAPT are obtained. From the consistency of the D3 and non-local results it seems safe to assume that the description of dispersion in both cases is quite reasonable, but the interaction energy is too strongly binding. The neglect of such terms raises the interaction energy and is assumed to accidentally improve the interaction energies. It is not clear whether this might be some fortunate error cancellation which can be found throughout the X23 set and could, therefore, be exploited. Hence, it seems appropriate to further observe the performance of PBE and PBE-D3.

Interestingly, a simple ac-QM/MM akin approach describing one molecule at the PBE-D3/TZVP level and the second one with point charges and LJ potentials provides significantly more positive (i.e. less binding) interaction energies. This is in stark disagreement with the previous findings of overbinding observed for the ac(1)-QM/MM scheme. This may very well be a hint that the negligence of higher order contributions due to the absence of a further point charge embedding does play an important role and should be considered. These results do not allow for a distinct decision on the validity of the supramolecular approach to interaction energies, but further testing with an embedding taken from an ac-QM/MM calculation is required.

Embedded Interaction Energies

For further tests a single-point ac(1)-QM/MM calculation of the cyanamide crystal was performed at the PBE-D3/TZVP+NPA+CGFF level of theory in order to generate an embedding environment. The unchanged environment was used for all monomer and dimer calculations. Only atomic charges and LJ potentials of the two monomers involved in each pair calculation were modified for the respective calculation.

Again, DFT-SAPT at the PBE/TZVP level of theory was used as the reference for interaction energies. These were calculated from a supramolecular approach, using PBE-D3/TZVP and PBE/TZVP, and at the QM/MM level. For the latter, the central monomer was calculated once with and once without atomic

Table 4.31: Gas phase interaction energies between pairs ($0+i$) of the central molecule 0 and the next molecules i (1 to 7) from the cyanamide crystal. DFT-SAPT calculations at the PBE/TZVP level of theory, supramolecular (Supra) interaction energy calculations at different levels of theory, and from an ac-QM/MM akin approach using PBE-D3/TZVP for the first molecule and point charges and LJ potentials for the second one (QM/MM). All values in $\text{kJ} \cdot \text{mol}^{-1}$.

	DFT-SAPT		Supra		QM/MM
	PBE/TZVP	PBE/TZVP	PBE-D3/TZVP	NL-PBE/TZVP	PBE-D3/TZVP
0+1	11.3	12.0	6.6	7.2	27.6
0+2	11.3	12.0	6.6	7.2	31.3
0+3	-16.5	-15.4	-20.6	-20.2	-8.2
0+4	-16.5	-15.4	-20.6	-20.2	-10.0
0+5	-3.0	-3.8	-7.3	-6.3	9.9
0+6	-26.4	-28.2	-32.4	-32.3	-21.3
0+7	-26.4	-28.2	-32.4	-32.3	-24.0
Sum	-66.3	-66.8	-100.1	-97.0	5.4

Table 4.32: Embedded interaction energies (ΔE) among pairs ($0+i$) of the central molecule (0) and the next molecules i (1 to 7) from the cyanamide crystal. DFT-SAPT calculations (DFT-SAPT) at the PBE/TZVP level, QM/MM calculations (QM/MM) at the PBE-D3/TZVP level of theory, supramolecular calculations (Supra) at the PBE/TZVP and PBE-D3/TZVP levels of theory. The point charge and LJ embedding from the a PBE-D3/TZVP+NPA+CGFF ac(1)-QM/MM calculation was used for all other calculations. All values in $\text{kJ} \cdot \text{mol}^{-1}$.

	DFT-SAPT		Supra		QM/MM
	PBE/TZVP	PBE/TZVP	PBE-D3/TZVP	PBE-D3/TZVP	PBE-D3/TZVP
0+1	25.7	26.6	21.2	14.4	14.4
0+2	25.7	26.6	21.2	17.6	17.6
0+3	-22.5	-21.4	-26.6	-21.8	-21.8
0+4	-22.5	-21.4	-26.6	-23.5	-23.5
0+5	-1.1	-1.2	-4.8	-4.0	-4.0
0+6	-35.1	-36.8	-41.1	-34.2	-34.2
0+7	-35.1	-36.8	-41.1	-35.7	-35.7
Sum	-64.8	-64.4	-97.7	-87.2	-87.2

charges and LJ potentials at the second monomer to determine the QM/MM interaction energy.

The results (Table 4.32) display a very consistent behaviour of the supramolecular interaction energies compared to the DFT-SAPT ones. Just as in the gas phase calculations a quite constant difference of about $5 \text{ kJ} \cdot \text{mol}^{-1}$ between supramolecular PBE-D3/TZVP and DFT-SAPT at the same level of theory is observed, yielding an RMSD of $4.8 \text{ kJ} \cdot \text{mol}^{-1}$. And again the supramolecular results without DFT-D3 corrections agree remarkably well with the DFT-SAPT results. A maximum difference of only $1.7 \text{ kJ} \cdot \text{mol}^{-1}$ and an RMSD of $1.1 \text{ kJ} \cdot \text{mol}^{-1}$ are obtained. Interestingly, the QM/MM calculations provide interaction energies much closer to DFT-SAPT than in the gas phase tests. With the exception of the first two pairs, the differences to DFT-SAPT are less than $3 \text{ kJ} \cdot \text{mol}^{-1}$, and the overall RMSD of $5.4 \text{ kJ} \cdot \text{mol}^{-1}$ is not much larger than for the supramolecular PBE-D3 results.

However, the ac(2)-QM/MM scheme uses combined interaction energy corrections rather than explicit interaction energies at the QM/MM and QM level. Therefore, it is tested whether these independently calculated interactions yield the same correction terms. The differences are found to be well below $0.1 \text{ kJ} \cdot \text{mol}^{-1}$. Thus, it can be assumed that the ac(2)-QM/MM method provides the same corrections as a stepwise subtraction and addition of QM/MM and QM contributions does.

Although it seems advisable to waive DFT-D3 corrections in order to gain presumably more accurate cohesive energies, it is still not performed routinely. For instance, the astonishing performance of QM/MM compared to the supramolecular approach with PBE-D3 indicates that the latter might just work remarkably bad for cyanamide. The better agreement of PBE without D3 might simply be fortunate error compensation. As first tests on the two-body corrections had not suggested general problems with PBE-D3 among the X23 molecules, a preference for the negligence of D3 was not made, but the performance of DFT with and without the correction was further tested. Moreover, systematic comparison with plane-wave DFT results

Table 4.33: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} with respect to experimental results. Values for ac(2)-QM/MM optimised structures at the Func/TZVP+NPA+CGFF (Func=PBE, PBE-D3, PBE-D3^{ABC}) levels of theory. ΔV in %, ΔE_{coh} in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV			ΔE_{coh}		
	PBE	PBE-D3	PBE-D3 ^{ABC}	PBE	PBE-D3	PBE-D3 ^{ABC}
1,4-cyclohexanedione	9.5	-3.1	-2.9	34.4	-16.3	-15.5
acetic acid	5.7	-4.0	-4.0	12.8	-14.4	-14.1
adamantane	15.8	-4.7	-4.5	52.1	-12.3	-10.3
ammonia	-8.2	-13.0	-13.1	-10.8	-25.0	-25.1
anthracene	17.7	-2.2	-2.1	82.5	-6.2	-3.9
benzene	15.2	-5.0	-4.9	30.4	-12.5	-12.1
carbon dioxide	7.5	-1.5	-1.5	8.0	-5.7	-5.7
cyanamide	2.9	-4.1	-4.2	-1.2	-26.6	-26.6
cytosine	11.9	-1.8	-2.1	45.8	-7.6	-7.0
ethylcarbamate	6.4	-4.5	-4.4	20.8	-15.6	-15.1
formamide	6.2	-4.5	-4.5	5.1	-17.7	-17.6
hexamine	12.3	-4.2	-4.0	45.6	-20.8	-19.0
imidazole	8.1	-2.1	-2.1	9.1	-22.3	-22.3
naphthalene	18.5	-2.1	-2.0	57.2	-9.3	-8.1
oxalic acid (α)	5.7	-2.8	-2.5	14.7	-20.1	-19.6
oxalic acid (β)	2.6	-3.3	-3.2	-9.8	-18.5	-18.2
pyrazine	10.2	-6.4	-6.4	25.9	-17.9	-17.8
pyrazole	9.8	-1.3	-1.3	18.3	-17.7	-17.4
succinic acid	9.5	-2.4	-2.4	30.0	-22.2	-21.4
triazine	5.4	-6.9	-6.9	27.3	-12.9	-12.9
trioxane	12.6	-0.2	-0.1	22.1	-17.3	-16.6
uracil	7.8	-3.5	-3.6	30.2	-21.5	-20.9
urea	2.4	-2.4	-2.4	-4.2	-31.4	-31.3
MD	8.5	-3.7	-3.7	23.8	-17.0	-16.5
MAD	9.2	3.7	3.7	26.0	17.0	16.5
RMSD	10.2	4.5	4.5	32.5	18.2	17.7

was considered a crucial test to assess the performance of the ac-QM/MM method, but from literature plane-wave results are only available with the D3 correction.

4.7.8 D3, or not D3, that is the Question

As the tests on cyanamide had suggested that PBE without D3 corrections might provide better interaction energies and, thus, cohesive energies, a series of calculations was performed on the X23 set. Several structures were fully optimised, including the optimisation of cell parameters, according to the ac(2)-QM/MM scheme. The PBE/TZVP+NPA+CGFF level of theory was used either with or without the inclusion of D3 corrections.

The resulting cell volume and cohesive energy differences to experiment in Table 4.33 show a distinctly different trend over the X23 set than the previous tests on cyanamide only. For all systems under investigation the optimised cell volumes are significantly larger if the D3 corrections are left out, with an RMSD of 10.2% among the X23 set. With the D3 corrections a significantly smaller RMSD of 4.5% is obtained. Especially for the non-polar molecules enormous volume deviations of up to 18.5% (naphthalene) are obtained without D3. This is in agreement with the expectation that dispersion interactions are dominating these structures. With PBE-D3 no such trend is observed, but the largest deviations are found for the hydrogen bonded system ammonia (-13.0%).

The same holds for the cohesive energies. These are always less negative (i.e. less binding) if D3 is not used, which directly corresponds to less stabilising intermolecular interactions. The MD of $23.8 \text{ kJ} \cdot \text{mol}^{-1}$ indicates a systematic underbinding without D3, although in few cases an overbinding is found as well. An overall RMSD of $32.5 \text{ kJ} \cdot \text{mol}^{-1}$ is obtained. Calculations with D3 provide a much smaller RMSD of $18.2 \text{ kJ} \cdot \text{mol}^{-1}$ which indicates a better agreement with the experimental results. The MD of $-17.0 \text{ kJ} \cdot \text{mol}^{-1}$ shows a systematic underbinding, a large discrepancy between the PBE and PBE-D3 results. This is, however, not

Table 4.34: Cell angle deviations $\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$, and their mean $M\Delta\angle$, cell volume deviation ΔV , and cohesive energy deviation ΔE_{coh} of unrestricted ac(2)-QM/MM optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory with respect to optimisations restricting all angles to their initial values. Angle differences in $^\circ$, ΔV in %, and ΔE_{coh} in $\text{kJ} \cdot \text{mol}^{-1}$.

	$\Delta\alpha$	$\Delta\beta$	$\Delta\gamma$	$M\Delta\angle$	ΔV	ΔE_{coh}
acetic acid	2.5	-0.2	-1.5	0.3	-1.6	-0.6
adamantane	0.0	0.0	0.0	0.0	-0.0	0.0
ammonia	-1.9	-1.9	-1.9	-1.9	0.3	-0.5
benzene	-0.9	-0.3	0.2	-0.4	-0.1	-0.1
carbon dioxide	-0.9	-0.9	-0.9	-0.9	-0.5	0.0
cyanamide	0.6	1.1	-0.9	0.3	0.1	-0.1
cytosine	0.5	-0.5	-0.9	-0.3	-0.2	-0.1
oxalic acid (α)	0.3	-0.4	3.3	1.1	-0.6	-0.7
pyrazine	2.2	0.0	0.0	0.7	-0.2	-0.4
pyrazole	0.8	4.6	0.4	1.9	0.5	-1.3
urea	0.0	0.0	-0.3	-0.1	0.0	0.0
MD	0.3	0.1	-0.2	0.1	-0.2	-0.3
MAD	1.0	0.9	0.9	0.9	0.4	0.3
RMSD	1.3	1.6	1.3	1.4	0.6	0.5

reflected in the values of the aforementioned cyanamide system. Without D3 in this case the smallest energy deviation to experiment is obtained among the whole X23 set, but the PBE-D3 results show the second largest deviations for this system. This leads to the conclusion that by chance for cyanamide PBE performs extraordinarily well, while PBE-D3 performs extraordinarily poorly. From these results there is no indication that the neglect of D3 corrections in general improves the results, hence, D3 corrections are used throughout all calculations.

The inclusion of three-body term within the D3 model (D3^{ABC}) does only have a minor effect on the results. Even for the largest system anthracene the difference in the cell volume decreases only slightly from -4.7 to -4.5% upon the change from PBE-D3 to PBE-D3^{ABC}. Among all systems the impact on the volumes is well below 0.1%. For the energies, a slight improvement of the RMSD is observed, but its change from 18.2 to 17.7 $\text{kJ} \cdot \text{mol}^{-1}$ is almost negligible. The largest change of 2.3 $\text{kJ} \cdot \text{mol}^{-1}$ (anthracene) represents an improvement of the energy, but still only a minor change. This is believed to be still negligible with respect to the expected accuracy of the method. Hence, the D3 scheme without three-body terms is used for all subsequent calculations. Nevertheless, the impact of D3^{ABC} terms will increase for larger molecules. Thus, for such calculations it needs to be reassessed if the use of D3^{ABC} can yield improved results.

It seems safe to conclude that the negligence of D3 corrections may sometimes lead to fortunate error cancellation, but generally does improve neither cell parameters nor cohesive energies. The inclusion of three-body terms in D3^{ABC} has only a marginal influence among the X23 set, although for the largest molecules small improvements can be gained.

4.7.9 Crystal Symmetry Restrictions

So far, the cell geometry has always been optimised without restriction. Although the observed changes in cell angles at the two-body level are small, the question remains how much this reduction of symmetry influences the results. As a test, all systems providing three square angles were optimised at the ac(2)-QM/MM PBE-D3/TZVP+NPA+CGFF level of theory with all angles either restricted or optimised. The final values in Table 4.34 show only very small changes in the angles, the overall RMSD is 1.3 $^\circ$. The respective volume changes are quite small as well, with differences mainly below 0.5% and an RMSD of 0.6%. Individual differences are partially positive and negative, thus, no trend towards larger or smaller cell volumes seems to be observable on the free optimisation of the angles. For the cohesive energies, the additional degrees of freedom regularly allow for slightly lower energies, but the differences are very small, with a RMSD of only 0.5 $\text{kJ} \cdot \text{mol}^{-1}$. Therefore, it is concluded that the error arising from breaking crystal symmetry is in fact negligible for the energy, albeit a small impact on the volumes can be observed. However, the freely

Table 4.35: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} with respect to experimental results from ac(1)-QM/MM (ac(1)) and ac(2)-QM/MM (ac(2)) optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory. ΔV in %, ΔE_{coh} in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV		ΔE_{coh}	
	ac(1)	ac(2)	ac(1)	ac(2)
1,4-cyclohexanedione	-1.2	-3.1	-6.4	-16.3
acetic acid	-5.0	-4.0	-21.2	-14.4
adamantane	-9.0	-4.7	-1.8	-12.3
ammonia	-14.4	-13.0	-5.9	-25.1
anthracene	-1.6	-2.2	-15.1	-6.2
benzene	-4.3	-5.0	-11.1	-12.5
carbon dioxide	-9.5	-1.5	-1.0	-5.7
cyanamide	-4.5	-4.1	-29.5	-26.6
cytosine	-5.5	-1.8	-62.8	-7.6
ethylcarbamate	-4.8	-4.5	-16.2	-15.6
formamide	-11.0	-4.5	-68.2	-20.4
hexamine	-3.2	-4.0	-4.1	-20.8
imidazole	-1.7	-2.1	-20.9	-22.3
naphthalene	-0.5	-2.1	-11.0	-9.3
oxalic acid (α)	-3.9	-2.8	-35.7	-20.1
oxalic acid (β)	-2.0	-3.3	-43.1	-18.5
pyrazine	-4.2	-6.4	-6.6	-17.8
pyrazole	3.8	-1.3	-7.1	-17.7
succinic acid	-6.7	-2.4	-61.5	-22.2
triazine	8.0	-6.9	14.8	-12.9
trioxane	-3.7	-0.2	-20.1	-17.3
uracil	-4.9	-3.5	-79.9	-21.5
urea	-14.1	-2.4	-90.0	-31.4
MD	-4.5	-3.7	-26.3	-17.2
MAD	5.5	3.7	27.6	17.2
RMSD	6.7	4.5	38.0	18.3

optimised angles may serve as an indicator to detect problems. Hence, it is deemed favourable to use the unrestricted optimisation. Moreover, this minor effect on both volumes and energies proves that it is a sound approach to rely on the implicit symmetry of the environment.

4.7.10 ac(2)-QM/MM Application

The basic ac(2)-QM/MM scheme, as described by Equation 4.65, was used to optimise the X23 structures at the PBE-D3/TZVP+NPA+CGFF level of theory. The selection of the two-body region was performed according to the atom distance approach with the predetermined cutoff distances r_R^f listed in Table 4.29.

The results in Table 4.35 allow for two main observations. Firstly, one- and two-body results qualitatively agree in the deviation to experiment. Both yield generally too small cell volumes and too strongly binding cohesive energies. Nevertheless, the statistics show that the deviations from experimental results are smaller with ac(2)-QM/MM than with ac(1)-QM/MM. For cell volumes the RMSD decreases from 6.7 to 4.5%, and for the cohesive energies it decreases from 38.0 to 18.3 $\text{kJ} \cdot \text{mol}^{-1}$. In fact, for ac(2)-QM/MM no positive cell volume deviations are found, which is a clear improvement to the few outliers existing for ac(1)-QM/MM. Moreover, very large energy deviations (e.g. $-90 \text{ kJ} \cdot \text{mol}^{-1}$ for urea) are not observed for ac(2)-QM/MM calculations, indicating more consistent results. Thus, a systematic improvement from the addition of two-body corrections can be verified.

4.7.11 ac(2-F)-QM/MM Application

Since analytical gradients are not available for the ac(2)-QM/MM scheme, only single-point energies were calculated. For all X23 structures ac(2-F)-QM/MM energies at the PBE-D3/TZVP+NPA+CGFF level of theory were computed at the ac(2)-QM/MM optimised geometries.

Table 4.36: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} with respect to experimental results from ac(2)-QM/MM (ac(2)) and ac(2-I)-QM/MM (ac(2-I)) optimisations at the PBE-D3/TZVP+NPA+CGFF level of theory. For the ac(2)-QM/MM optimised geometry, the ac(2-F)-QM/MM energy was calculated as well (ac(2-F)). ΔV in %, ΔE_{coh} in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV		ΔE_{coh}		
	ac(2)	ac(2-I)	ac(2)	ac(2-F)	ac(2-I)
acetic acid	-4.8	-1.1	-14.1	1.1	-1.9
adamantane	-4.7	-4.6	-12.3	-12.0	-12.1
ammonia	-10.4	-8.5	-20.6	-16.0	-17.4
anthracene	-2.3	-0.3	-4.7	6.8	6.1
benzene	-5.0	-3.1	-12.6	-8.0	-8.1
carbon dioxide	-1.5	0.0	-5.7	-4.4	-4.7
cyanamide	-6.0	-5.8	-31.0	-22.8	-25.9
cytosine	-2.9	0.3	-26.4	8.0	-9.1
ethylcarbamate	-5.3	-2.6	-17.0	-8.8	-9.4
formamide	-4.8	5.4	-17.5	1.5	-0.4
hexamine	-4.8	-3.9	-18.4	-14.8	-14.5
imidazole	-3.9	4.8	-20.1	-0.1	-7.4
naphthalene	-2.0	0.2	-6.5	1.5	1.1
oxalic acid (α)	-2.7	1.5	-20.1	10.8	0.9
oxalic acid (β)	-4.0	0.8	-22.8	14.5	2.8
pyrazine	-8.4	2.8	-16.2	-10.3	-14.3
pyrazole	-0.3	-0.9	-15.5	-2.6	-11.1
succinic acid	-3.4	-1.2	-63.4	-58.3	-58.3
triazine	-6.0	-4.7	-7.8	-4.9	-5.0
trioxane	0.5	3.2	-12.6	-6.2	-5.2
uracil	-4.7	-2.7	-56.3	-44.1	-43.9
urea	-2.4	5.1	-31.4	-4.8	-6.4
MD	-4.1	-0.7	-20.6	-7.9	-11.1
MAD	4.1	2.9	20.6	11.9	12.1
RMSD	4.7	3.6	25.1	18.2	18.4

The results in Table 4.36 are somewhat mixed. Comparison of the ac(2-F)-QM/MM energies calculated from ac(2)-QM/MM optimised structures shows a very consistent trend. All cohesive energies are less binding for ac(2-F)-QM/MM than for ac(2)-QM/MM, although in some cases the changes are small. Due to the overbinding of ac(2)-QM/MM energies with respect to experimental results, the less binding ac(2-F)-QM/MM energies are regularly closer to experiment. In 6 cases the shift even results in underbinding up to $14.5 \text{ kJ} \cdot \text{mol}^{-1}$ (oxalic acid (β)) for ac(2-F)-QM/MM. Consequently, the mean deviation decreases from -20.6 to $-7.9 \text{ kJ} \cdot \text{mol}^{-1}$ from ac(2)-QM/MM to ac(2-F)-QM/MM, and the RMSD decreases notably from 25.1 to $18.2 \text{ kJ} \cdot \text{mol}^{-1}$. Albeit this seems to be a reasonable improvement, the lack of analytical gradients limits its usefulness.

4.7.12 ac(2-I)-QM/MM Application

In contrast to the ac(2-F)-QM/MM approach, ac(2-I)-QM/MM analytical gradients are available. The approach was applied for optimisations of the X23 molecules at the PBE-D3/TZVP+NPA+CGFF level of theory.

The results in Table 4.36 show a systematic trend in both cell volumes and cohesive energies. Volumes are always larger for ac(2-I)-QM/MM than for ac(2)-QM/MM, with the only exception of pyrazole for which a 0.6% smaller volume is obtained. Hence, the volume differences with respect to experiment tend to become smaller. In several cases, ac(2-I)-QM/MM volumes even become larger than experimental ones, therefore, an overall mean deviation of only -0.7% is obtained (-4.1% for ac(2)-QM/MM). However, the differences to experiment do not change uniformly, e.g. for adamantane deviations of -4.6 and -4.7% are obtained for ac(2)-QM/MM and ac(2-I)-QM/MM, but for pyrazine the change is from -8.4 to 2.8% . Consequently, the RMSD with respect to experiment becomes a bit smaller for ac(2-I)-QM/MM than for ac(2)-QM/MM (4.7 and 3.6%), but much less than expected from the MD. Moreover, the ac(2)-QM/MM results include

the expected underestimation of cell volumes. Hence, it is not clear whether the smaller RMSD for ac(2-I)-QM/MM volumes indicates more reasonable results.

In the cohesive energies a similar trend as for ac(2-F)-QM/MM is found. All energies are less binding than the ac(2)-QM/MM ones. The changes compared to ac(2)-QM/MM vary from 0.2 to 25.6 kJ · mol⁻¹. As for ac(2-F)-QM/MM the shift to less binding cohesive energies yields smaller differences to experiment, in some cases even an underbinding of ac(2-I)-QM/MM is observed. Albeit the energies are shifted up, the spread in the differences remains approximately the same. The differences span a range of about 60 kJ · mol⁻¹, from -56.3 to 6.1 kJ · mol⁻¹. Consequently, the MD is almost halved from -20.6 to -11.1 kJ · mol⁻¹, and the RMSD is still significantly reduced from 25.1 to 18.4 kJ · mol⁻¹. Although some deviations in the volume are to the wrong direction, the energies seem to be systematically improved with respect to the ac(2)-QM/MM scheme.

The demand in computational time, however, is significantly larger than for ac(2)-QM/MM, as one additional monomer calculation (for $E_{\text{emb}[0]}(m)$) is required. Nonetheless, for large systems the cost of the dimer calculations will become the limiting factor. This is the same in both ac(2-I)-QM/MM and ac(2)-QM/MM, therefore, these should be equally applicable. Further benchmarking calculations on both ac(2)-QM/MM and ac(2-I)-QM/MM were carried out to better compare their reliability and accuracy. These are presented in Section 5.1.2.

4.8 Suggested Models for ac-QM/MM

After the ac-QM/MM internal parameters are (more or less) optimised, the performance with different external variables needs to be evaluated to some extent. Four main parameters were expected to show a strong influence on the results, namely the QM method, the basis set, the population analysis, and the definitions of LJ parameters.

4.8.1 ac(1)-QM/MM

Parameter Selection

At the ac(1)-QM/MM level, the basis set is expected to have only a minor effect. Preliminary test with the default setup (PBE-D3 functional, NPA charges, CGFF parameters) had also provided negligibly small differences in all parameters (typically less than 0.1% in V and less than 1 kJ · mol⁻¹ in E_{coh}). Thus, the functional was varied (PBE-D3 and TPSS-D3), while the basis set was always TZVP. Hybrid DFT functionals were not considered, since the reference values⁹⁸ from plane-wave DFT optimisations of the X23 set are limited to (meta-)GGA functionals.

Albeit NPA charges are known to be reasonable with respect to chemical intuition and stable with respect to the basis size, they do not correctly describe the molecular multipoles. For that reason, the Chelpg model was tested as well. Since these charges are fitted to represent the QM electrostatic potential around the molecule, they should in principle provide a more reasonable electrostatic embedding than NPA charges do. At the ac(1)-QM/MM level, intermolecular interactions are completely described by QM/MM, thus, the choice of LJ parameters is crucial for the results. Four force fields with the respective parameter sets were taken into account, namely UFF, GAFF, CGFF and MM3. The initial tests already revealed one critical demand for the LJ parameters. All atoms need to have LJ parameters, otherwise there is no barrier to prevent a positively charged MM atom move towards the QM electron density. The MM atom may, therefore, come unphysically close to a QM nucleus, leading to unreasonable structures. This regularly happened for systems featuring intermolecular hydrogen bonds, since in that cases some MM hydrogens are positively charged and close to the QM part. As GAFF, in contrast to the other force fields, does not include a LJ parameter for hydrogen, GAFF was excluded from further calculations. Although one LJ parameter for hydrogen is sufficient to prevent the aforementioned unphysical behaviour, it seems that at least two different LJ parameters for hydrogen atoms are required to obtain reasonable results. Since hydrogen bonds are usually significantly shorter than non-bonding hydrogen contacts, a single LJ parameter for all hydrogens is most likely fitted to describe the latter ones better than the former ones. Hence, for hydrogen bonded systems geometry optimisations result in unreasonably long intermolecular hydrogen bonds. In the case of oxalic acid (β), for example, hydrogen bond lengths of 2.50 Å are obtained with UFF, while more reasonable values of 1.60 Å for CGFF (and 1.54 Å from a two-body ac-QM/MM calculation) can be observed. Table 4.37 shows the influence on cell parameters for this system as well as for an uncritical case, benzene. While for the latter quite reasonable parameters are obtained, the first one shows enormous differences. For oxalic acid (β) the cell volume is overestimated by almost 16%, much more than for any calculation with CGFF or MM3 force fields. The main elongation takes place along the \vec{a} cell direction to which the intermolecular

Table 4.37: Cell parameters and cohesive energies for one hydrogen bonded system, oxalic acid (β), and one vdW bonded system, benzene. ac(1)-QM/MM PBE-D3/TZVP+NPA+UFF optimised (Opt.) and experimental (Exp.) results. Cell lengths (a, b, c) in Å, angles (α, β, γ) in °, volumes V in Å³, and cohesive energies E_{coh} in kJ · mol⁻¹.

		a	b	c	α	β	γ	V	E_{coh}
benzene	Exp.	7.390	9.420	6.810	90.0	90.0	90.0	474.1	-51.9
	Opt.	7.532	9.357	6.849	89.2	89.3	90.0	482.6	-70.3
oxalic acid (β)	Exp.	5.330	6.015	5.436	90.0	115.8	90.0	156.9	-96.3
	Opt.	6.058	6.385	5.631	91.3	123.6	88.5	181.4	-56.1

Table 4.38: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} between PBE-D3 and TPSS-D3 from ac(1)-QM/MM optimisations with various other theory parameter. ΔV in % and ΔE_{coh} in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV				ΔE_{coh}			
	NPA		Chelpg		NPA		Chelpg	
	CGFF	MM3	CGFF	MM3	CGFF	MM3	CGFF	MM3
1,4-cyclohexanedione	0.1	0.3	0.0	-0.1	1.5	1.5	1.3	1.4
acetic acid	0.2	0.1	0.5	0.4	0.3	0.8	4.6	1.7
adamantane	-4.5	0.3	0.2	0.3	0.2	0.1	0.1	0.1
ammonia	-0.5	-0.9	-0.2	-0.3	-2.1	-1.4	-2.0	-1.1
anthracene	-0.0	-0.1	0.1	0.2	-2.9	-2.8	-0.4	-0.4
benzene	-0.0	-0.1	0.3	0.1	-1.6	-1.4	-0.3	-0.3
carbon dioxide	0.4	0.8	0.1	0.8	0.7	0.5	0.3	0.2
cyanamide	-0.2	1.0	0.1	-0.1	-2.4	2.9	-2.8	2.1
cytosine	0.1	-0.4	0.1	-0.3	-0.3	0.1	-0.6	0.7
ethylcarbamate	-0.1	0.7	-0.1	0.4	0.0	0.4	-0.1	0.6
formamide	0.1	0.6	0.4	0.4	0.5	0.4	3.2	0.2
hexamine	0.1	-0.1	0.0	0.1	-1.2	-0.7	0.4	0.4
imidazole	-0.2	0.2	-1.1	0.2	-0.6	0.6	0.2	-0.3
naphthalene	0.0	-0.1	0.2	0.1	-2.1	-2.0	-0.4	-0.4
oxalic acid (α)	0.3	0.5	0.3	2.2	2.6	0.7	2.6	16.8
oxalic acid (β)	0.3	3.9	0.3	-0.1	1.8	0.4	1.4	2.6
pyrazine	0.1	-0.1	0.0	0.2	-1.2	-0.8	-0.1	0.1
pyrazole	-0.2	-0.0	0.1	0.1	-1.4	-1.3	-0.5	-0.2
succinic acid	0.1	0.2	0.2	-0.3	1.0	1.7	2.0	3.1
triazine	0.1	0.1	0.4	0.4	-0.1	-0.2	0.4	0.5
trioxane	0.1	0.2	0.2	0.3	1.6	1.3	2.6	1.9
uracil	-1.7	0.2	0.1	2.6	2.2	0.2	0.5	-15.7
urea	0.1	0.2	0.2	0.2	-0.7	-0.2	-0.6	0.1
MD	-0.2	0.3	0.1	0.3	-0.3	-0.0	0.4	1.2
RMSD	1.0	0.9	0.3	0.8	1.5	1.2	1.7	4.9

hydrogen bonds are aligned. Consequently, the corresponding interactions are heavily underestimated, and the cohesive energy is by $40 \text{ kJ} \cdot \text{mol}^{-1}$ underbinding. Since the overestimation of electrostatic interactions should lead to overbinding (as it can regularly be observed for both CGFF and MM3), this does not encourage the use of UFF. The expectable errors in both geometry and energy from UFF for hydrogen bonded systems were deemed unacceptable and UFF was excluded from the set of force fields for further tests.

Thus, the final set of parameters to be tested consisted of the functionals PBE-D3 and TPSS-D3, the population schemes NPA and Chelpg, and the force fields CGFF and MM3. All eight combinations of these were used throughout the X23 set to perform a full ac(1)-QM/MM optimisation of both geometry and cell parameters. The results for cell volumes and cohesive energies reveals a series of trends, best evaluated individually for each variation parameter. Each pair of parameters to be compared, e.g. the functionals PBE-D3 and TPSS-D3, is combined with all four combinations of residual parameters, e.g. NPA+CGFF, NPA+MM3, Chelpg+CGFF, Chelpg+MM3. Therefore, one RMSD value results from the comparison between PBE-D3 and TPSS-D3 with each of the residual parameter combinations.

Influence of the Functional

As it had been expected from the fact that intermolecular interactions are solely described at the QM/MM level, the choice of the functional does not have a notable influence on V and E_{coh} (see Table 4.38). Individual differences between PBE-D3 and TPSS-D3 for the same set of residual parameters are mostly well below 1%, although some outliers are observed, namely adamantane in conjunction with NPA+CGFF (4.5%), oxalic acid (α) with Chelpg+MM3 (2.2%), oxalic acid (β) with NPA+MM3 (3.9%) and uracil with either NPA+CGFF (1.7%) or Chelpg+MM3 (2.6%). The differences in E_{coh} look very similar. For almost all systems the difference between PBE-D3 and TPSS-D3 is only up to $4 \text{ kJ} \cdot \text{mol}^{-1}$, which is very comparable to the differences between PBE-D3 and TPSS-D3 presented for plane-wave calculations with the same functionals.⁹⁸ Two distinct outliers are found with Chelpg+MM3, namely oxalic acid (α) and uracil,

Table 4.39: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} between NPA and Chelpg results from ac(1)-QM/MM optimisations with various other theory parameter. ΔV in % and ΔE_{coh} in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV				ΔE_{coh}			
	PBE-D3		TPSS-D3		PBE-D3		TPSS-D3	
	CGFF	MM3	CGFF	MM3	CGFF	MM3	CGFF	MM3
1,4-cyclohexanedione	0.7	1.2	0.6	0.8	2.9	3.3	2.7	3.2
acetic acid	1.3	2.1	1.6	2.5	11.2	14.3	15.5	15.2
adamantane	-4.7	-0.1	-0.0	-0.1	-0.4	-0.4	-0.5	-0.4
ammonia	6.1	8.1	6.4	8.8	19.0	11.3	19.1	11.6
anthracene	-1.5	-2.4	-1.3	-2.2	-21.5	-21.4	-19.0	-19.0
benzene	-2.3	-3.3	-2.0	-3.1	-11.4	-10.4	-10.1	-9.3
carbon dioxide	1.1	1.6	0.8	1.6	2.1	1.6	1.7	1.3
cyanamide	5.0	6.1	5.2	5.0	59.9	28.8	59.5	28.0
cytosine	0.5	1.0	0.6	1.1	23.8	4.2	23.5	4.8
ethylcarbamate	0.7	-0.0	0.7	-0.3	3.6	0.5	3.5	0.7
formamide	-0.7	-0.6	-0.5	-0.9	1.8	-2.8	4.5	-3.0
hexamine	0.1	-0.0	0.1	0.2	-0.6	-0.5	1.0	0.6
imidazole	1.3	0.2	0.5	0.3	3.2	-0.1	4.0	-1.0
naphthalene	-0.8	-2.8	-0.6	-2.5	-14.2	-14.5	-12.5	-12.9
oxalic acid (α)	2.5	4.6	2.4	6.4	28.8	25.9	28.8	42.0
oxalic acid (β)	1.7	7.1	1.7	2.9	12.6	28.9	12.2	31.1
pyrazine	0.7	-1.0	0.7	-0.7	-13.6	-11.5	-12.5	-10.6
pyrazole	-1.6	-0.8	-1.3	-0.7	0.5	-2.8	1.4	-1.7
succinic acid	0.6	0.9	0.7	0.4	7.2	7.9	8.2	9.3
triazine	-0.4	-0.0	-0.1	0.3	-1.4	-1.2	-0.9	-0.5
trioxane	-1.7	-2.9	-1.6	-2.7	-15.9	-13.9	-14.9	-13.3
uracil	-1.6	-2.9	0.2	-0.6	-3.9	8.5	-5.6	-7.4
urea	1.8	1.3	1.8	1.4	21.5	7.6	21.6	7.9
MD	0.4	0.8	0.7	0.8	4.9	2.7	5.7	3.9
RMSD	2.3	3.2	2.0	2.9	18.0	13.2	17.8	14.8

for which much larger differences of 16.8 and $15.7 \text{ kJ} \cdot \text{mol}^{-1}$ are found. These two values seem unreliable, although the structures are very similar (RMSDs of 0.003 and 0.015 \AA for oxalic acid (α) and uracil). The large energy differences stem from the LJ contribution (oxalic acid (α)) or from the electrostatic contribution (uracil), as the final atomic charges between PBE-D3 and TPSS-D3 differ by up to 0.2 e for some atoms. Both indicate that the combination of Chelpg with MM3 tends to be very sensitive to small structural changes and might, therefore, be less suited for the ac(1)-QM/MM scheme. Moreover, the comparison of NPA charges between the final uracil structures of the analogous optimisations shows charge differences of less than 0.02 e . This supports the impression that Chelpg charges are less reliable than NPA ones. Nevertheless, RMSD values for the parameter combinations NPA+CGFF, NPA+MM3, Chelpg+CGFF, and Chelpg+MM3 are still quite small. In V values of only 1.0 , 0.9 , 0.3 , and 0.8% are obtained, in E_{coh} values of 1.5 , 1.2 , 1.7 , and $4.9 \text{ kJ} \cdot \text{mol}^{-1}$. The latter value is notably higher, due to the aforementioned outliers, but calculation of RMSDs over all systems and parameter combinations yield only 0.8% in V and $2.8 \text{ kJ} \cdot \text{mol}^{-1}$ in E_{coh} , which does imply a very similar behaviour of both functionals PBE-D3 and TPSS-D3. Therefore, the computationally less costly PBE-D3 was selected as the default for most of the further studies.

Influence of the Population Analysis

The comparison of results between NPA and Chelpg charges (Table 4.39) shows much larger differences both in V and E_{coh} , but no clear trends. Differences in V are up to 8.8% (ammonia with TPSS-D3+MM3), and RMSDs of 2.3 , 3.2 , 2.0 , and 2.9% for the parameter combinations PBE-D3+CGFF, PBE-D3+MM3, TPSS-D3+CGFF, and TPSS-D3+MM3 are obtained. In E_{coh} differences up to $59.9 \text{ kJ} \cdot \text{mol}^{-1}$ (cyanamide with PBE-D3+CGFF) and RMSDs of 18.0 , 13.2 , 17.8 , and $14.8 \text{ kJ} \cdot \text{mol}^{-1}$ are found for these combinations. These results, as well as the RMSD calculated over all systems and parameter combinations (2.6% in V and $16.1 \text{ kJ} \cdot \text{mol}^{-1}$ in E_{coh}) show much more influence of the population analysis than of the functional. However, deviations take place in both directions for V as well as for E_{coh} . It seems that on average the

Table 4.40: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} between CGFF and MM3 results from ac(1)-QM/MM optimisations with various other theory parameter. ΔV in % and ΔE_{coh} in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV				ΔE_{coh}			
	PBE-D3		TPSS-D3		PBE-D3		TPSS-D3	
	CGFF	MM3	CGFF	MM3	CGFF	MM3	CGFF	MM3
1,4-cyclohexanedione	-8.8	-8.4	-8.7	-8.5	-21.1	-20.7	-21.1	-20.6
acetic acid	-4.7	-3.9	-4.8	-4.0	-15.7	-12.6	-15.2	-15.5
adamantane	-17.5	-13.6	-13.4	-13.5	-22.2	-22.2	-22.3	-22.2
ammonia	-19.0	-17.5	-19.4	-17.6	-11.0	-18.7	-10.3	-17.8
anthracene	-3.1	-4.0	-3.1	-4.0	-14.2	-14.1	-14.1	-14.1
benzene	-5.6	-6.5	-5.7	-6.8	-8.7	-7.7	-8.5	-7.7
carbon dioxide	-12.6	-12.1	-12.2	-11.5	-11.6	-12.1	-11.8	-12.2
cyanamide	-14.0	-13.1	-13.0	-13.2	-45.5	-76.6	-40.2	-71.7
cytosine	-9.1	-8.7	-9.5	-9.0	-90.6	-110.2	-90.2	-108.9
ethylcarbamate	-11.8	-12.5	-11.2	-12.1	-32.2	-35.3	-31.8	-34.6
formamide	-18.8	-18.7	-18.4	-18.7	-75.6	-80.2	-75.7	-83.2
hexamine	-6.5	-6.6	-6.6	-6.5	-16.5	-16.4	-16.0	-16.4
imidazole	-6.5	-7.5	-6.1	-6.3	-28.8	-32.1	-27.6	-32.6
naphthalene	-3.6	-5.5	-3.7	-5.6	-11.0	-11.3	-10.9	-11.3
oxalic acid (α)	-0.2	1.9	-0.0	3.8	-21.1	-24.0	-23.0	-9.8
oxalic acid (β)	-4.4	0.8	-0.9	0.3	-24.5	-8.2	-25.9	-7.0
pyrazine	-6.3	-7.9	-6.5	-7.7	-11.9	-9.8	-11.5	-9.6
pyrazole	-0.6	0.3	-0.3	0.3	-11.4	-14.7	-11.3	-14.4
succinic acid	-7.1	-6.8	-7.0	-7.3	-40.1	-39.4	-39.4	-38.3
triazine	11.9	12.4	12.0	12.4	0.1	0.3	0.0	0.4
trioxane	-9.8	-10.9	-9.7	-10.8	-20.8	-18.8	-21.1	-19.5
uracil	-10.0	-11.1	-8.2	-9.0	-89.0	-76.6	-91.0	-92.8
urea	-26.4	-26.8	-26.4	-26.7	-93.8	-107.7	-93.3	-107.0
MD	-8.4	-8.1	-8.0	-7.9	-22.3	-24.6	-22.0	-23.8
RMSD	11.4	11.2	10.9	11.1	41.7	46.3	41.6	47.0

cohesive energies tend to be more negative (i.e. stronger binding) with Chelpg charges than with NPA (by $4.3 \text{ kJ} \cdot \text{mol}^{-1}$ on average), but the difference is small compared to the RMSD of $16.1 \text{ kJ} \cdot \text{mol}^{-1}$ between NPA and Chelpg. Thus, no general trend for the influence of the charges can be derived. Both NPA and Chelpg seem to be reasonable choices for the electrostatic embedding and no strong preference is based on these results.

Influence of the Force Field

Even larger differences are obtained from the comparison of CGFF and MM3 results (Table 4.40). Almost all CGFF volumes are significantly smaller than the MM3 results, but some exceptions are observed. For oxalic acid (α), oxalic acid (β), and pyrazole both results are very similar. Dependent on the other parameters CGFF volumes are slightly smaller or larger with respect to MM3. Only triazine shows the opposite behaviour, in that case V is by around 12% larger with CGFF than with MM3 for each combination of residual parameters. Very large deviations up to 26.7% (urea with TPSS+Chelpg) are obtained, but the behaviour is quite similar for all combinations of the residual parameters. For PBE-D3+NPA, PBE-D3+Chelpg, TPSS-D3+NPA, and TPSS-D3+Chelpg very similar volume RMSDs over the X23 systems are obtained (11.4, 11.2, 10.9, and 11.1%), yielding an RMSD of 11.2% in V over all combinations. These results show a crucial influence of the LJ parameters used. A strong impact on the cohesive energies can be observed as well. For all systems, again with the exception of triazine, the cohesive energies are more negative (i.e. stronger binding) with CGFF than with MM3. The deviations range from less than 10.0 up to $110.2 \text{ kJ} \cdot \text{mol}^{-1}$ (cytosine in conjunction with PBE-D3+NPA) among the benchmark set. Consequently, large RMSD values of 41.7, 46.3, 41.6, and $47.0 \text{ kJ} \cdot \text{mol}^{-1}$ are observed for the combinations PBE-D3+NPA, PBE-D3+Chelpg, TPSS-D3+NPA, and TPSS-D3+Chelpg. Over all combinations a RMSD of $44.2 \text{ kJ} \cdot \text{mol}^{-1}$ is obtained, which is much larger than the respective values from the comparison between either PBE and TPSS or NPA and Chelpg. Compared to CGFF, MM3 in almost all cases results in significantly large cells and less

Table 4.41: RMSD of molecular geometry without hydrogen atoms between various ac(1)-QM/MM optimised structures and the experimental results. All values in Å.

	PBE-D3/TZVP				TPSS-D3/TZVP			
	NPA		Chelpg		NPA		Chelpg	
	CGFF	MM3	CGFF	MM3	CGFF	MM3	CGFF	MM3
1,4-cyclohexanedione	0.046	0.048	0.023	0.024	0.038	0.038	0.019	0.019
acetic acid	0.018	0.018	0.019	0.019	0.019	0.019	0.019	0.020
adamantane	0.009	0.010	0.009	0.010	0.010	0.011	0.011	0.011
ammonia	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
anthracene	0.010	0.010	0.010	0.010	0.008	0.009	0.008	0.009
benzene	0.008	0.008	0.006	0.007	0.007	0.007	0.006	0.006
carbon dioxide	0.014	0.014	0.014	0.014	0.012	0.013	0.012	0.012
cyanamide	0.010	0.016	0.012	0.021	0.009	0.018	0.011	0.023
cytosine	0.021	0.020	0.021	0.027	0.022	0.021	0.021	0.028
ethylcarbamate	0.029	0.031	0.037	0.032	0.024	0.031	0.027	0.030
formamide	0.015	0.015	0.014	0.014	0.015	0.014	0.014	0.014
hexamine	0.012	0.012	0.012	0.012	0.015	0.015	0.015	0.016
imidazole	0.015	0.015	0.015	0.011	0.015	0.015	0.014	0.012
naphthalene	0.004	0.004	0.004	0.004	0.004	0.003	0.004	0.004
oxalic acid (α)	0.018	0.017	0.021	0.019	0.019	0.017	0.022	0.021
oxalic acid (β)	0.022	0.021	0.023	0.021	0.023	0.023	0.023	0.021
pyrazine	0.008	0.009	0.009	0.009	0.007	0.008	0.008	0.008
pyrazole	0.015	0.015	0.015	0.015	0.014	0.015	0.014	0.015
succinic acid	0.025	0.026	0.030	0.033	0.025	0.027	0.032	0.038
triazine	0.025	0.025	0.027	0.027	0.025	0.025	0.026	0.027
trioxane	0.015	0.015	0.015	0.015	0.015	0.014	0.014	0.014
uracil	0.033	0.025	0.024	0.028	0.022	0.025	0.024	0.025
urea	0.020	0.012	0.021	0.010	0.021	0.012	0.022	0.010
MD	0.017	0.017	0.017	0.017	0.016	0.017	0.016	0.017
RMSD	0.020	0.019	0.019	0.019	0.018	0.019	0.018	0.019

binding cohesive energies. The latter might possibly be a result of electrostatic interactions being reduced if intermolecular distances are enlarged in larger cells.

Comparison to Experimental Data

Further analysis is based on the comparison with experimental results in order to decide on the most appropriate set of parameters for ac(1)-QM/MM calculations. The RMSD on molecular geometries without hydrogen (Table 4.41), percentage cell volume differences (Table 4.42) and cohesive energy differences (Table 4.43) all refer to differences of calculated values with respect to the experimental values taken from literature.^{96,97} Interestingly, the internal geometry seems to be quite insensitive with respect to the theory parameters under investigation. Although in some cases the RMSD values are significantly different, e.g. for 1,4-cyclohexanedione these range from 0.0481 Å (PBE-D3/TZVP+NPA+MM3) to 0.0189 Å (TPSS-D3/TZVP+Chelpg+MM3), they are mostly quite similar among the different calculations of the same system. Taking all X23 systems into account overall RMSDs are nearly the same for PBE-D3/TZVP (0.0198, 0.0194, 0.0186, and 0.0188 Å for the combinations NPA+CGFF, NPA+MM3, Chelpg+CGFF, and Chelpg+MM) as well as for TPSS-D3/TZVP (0.0181, 0.0188, 0.0178, and 0.0189 Å). Even the difference between the functionals, which should have the largest influence on the internal geometry, is very small and further supports the impression that TPSS-D3 does not provide a notable improvement compared to PBE-D3.

The cell volume differences (Table 4.42) already show a clear distinction between the results with CGFF and those with MM3. With CGFF volumes are systematically smaller than the experimental values. Only in the cases of pyrazole (3.8, 5.5, 4.0%, and 5.4% with PBE-D3+NPA, PBE-D3+Chelpg, TPSS-D3+NPA, and TPSS-D3+Chelpg) and triazine (8.0, 8.5, 7.9, and 8.0%) larger volumes are obtained. Across the X23 set the MD for CGFF is -4.6% (PBE-D3+NPA) or -4.9% (PBE-D3+Chelpg). Very similar values of -4.4% (TPSS-D3+NPA) and -5.0% (TPSS-D3+Chelpg) are obtained with TPSS. In contrast, MM3 provides overall too

Table 4.42: Percentage cell volume deviation between ac(1)-QM/MM optimised structures at various levels of theory and the experimental results.

	PBE-D3/TZVP				TPSS-D3/TZVP			
	NPA		Chelpg		NPA		Chelpg	
	CGFF	MM3	CGFF	MM3	CGFF	MM3	CGFF	MM3
1,4-cyclohexanedione	-1.2	8.4	-1.9	7.1	-1.3	8.1	-1.9	7.2
acetic acid	-5.0	-0.3	-6.2	-2.4	-5.2	-0.4	-6.7	-2.8
adamantane	-9.0	10.3	-4.6	10.5	-4.8	10.0	-4.8	10.2
ammonia	-14.4	5.7	-19.3	-2.2	-14.0	6.6	-19.2	-1.9
anthracene	-1.6	1.5	-0.1	4.0	-1.6	1.6	-0.3	3.9
benzene	-4.3	1.4	-2.1	4.8	-4.3	1.4	-2.4	4.7
carbon dioxide	-9.5	3.5	-10.5	1.8	-9.9	2.6	-10.6	1.1
cyanamide	-4.5	11.0	-9.0	4.6	-4.4	9.9	-9.1	4.7
cytosine	-5.5	4.0	-6.0	3.0	-5.5	4.4	-6.0	3.3
ethylcarbamate	-4.8	8.0	-5.5	8.0	-4.7	7.3	-5.4	7.6
formamide	-11.0	9.6	-10.4	10.2	-11.1	8.9	-10.7	9.8
hexamine	-3.2	3.5	-3.2	3.6	-3.2	3.6	-3.3	3.5
imidazole	-1.7	5.1	-3.0	4.8	-1.5	4.9	-2.0	4.6
naphthalene	-0.5	3.2	0.3	6.2	-0.5	3.3	0.1	6.0
oxalic acid (α)	-3.9	-3.8	-6.3	-8.0	-4.2	-4.2	-6.5	-10.0
oxalic acid (β)	-2.0	2.5	-3.6	-4.3	-2.3	-1.4	-3.9	-4.2
pyrazine	-4.2	2.2	-4.9	3.3	-4.3	2.4	-4.9	3.0
pyrazole	3.8	4.3	5.5	5.2	4.0	4.4	5.4	5.0
succinic acid	-6.7	0.4	-7.3	-0.5	-6.8	0.2	-7.4	-0.2
triazine	8.0	-3.5	8.5	-3.5	7.9	-3.6	8.0	-3.9
trioxane	-3.7	6.8	-2.0	9.9	-3.8	6.5	-2.3	9.5
uracil	-6.7	3.6	-5.2	6.7	-5.1	3.4	-5.3	4.0
urea	-14.1	16.7	-15.6	15.2	-14.2	16.5	-15.7	15.0
MD	-4.6	4.5	-4.9	3.8	-4.4	4.2	-5.0	3.5
RMSD	6.8	6.5	7.6	6.6	6.5	6.3	7.7	6.5

large cell volumes, with the exception of acetic acid (-0.3 and -2.4%), oxalic acid (α) (-3.8 and -8.0%), and triazine (-3.5 and -3.5%). MM3 yields MDs of 4.5% (PBE-D3+NPA), 3.8% (PBE-D3+Chelpg), 4.2% (TPSS-D3+NPA), and 3.5% (TPSS-D3+Chelpg) to experimental results. All RMSD values are quite close together, ranging from 6.5% (PBE-D3+NPA+MM3) to 7.7% (TPSS+Chelpg+CGFF). Although the absolute deviations are comparably large for CGFF and MM3, the CGFF ones are deemed more reasonable. As no thermal effects are considered, the calculated values should rather be smaller than the experimental ones measured at finite temperature, consequently, the overall CGFF deviation is considered to be in the correct direction.

A somewhat similar behaviour is observed for the cohesive energy deviations with CGFF parameters. All differences are negative (i.e. stronger binding) for CGFF, with the exception of triazine (14.9 , 16.3 , 15.0 , and 15.9 $\text{kJ} \cdot \text{mol}^{-1}$ for PBE-D3+NPA, PBE-D3+Chelpg, TPSS-D3+NPA, and TPSS-D3+Chelpg). Very large values up to -111.5 $\text{kJ} \cdot \text{mol}^{-1}$ (urea with PBE-D3+Chelpg+CGFF) are obtained. Large RMSD values of 37.9 and 48.2 $\text{kJ} \cdot \text{mol}^{-1}$ (PBE-D3+NPA and PBE-D3+Chelpg) are found, showing a systematic strong overbinding throughout the X23 set. The MM3 results, on the other hand, show a quite broad range of deviations between -57.3 $\text{kJ} \cdot \text{mol}^{-1}$ (oxalic acid (α) with TPSS-D3+Chelpg) and 27.8 $\text{kJ} \cdot \text{mol}^{-1}$ (cytosine with PBE-D3+NPA), and rather small MDs of 4.7 $\text{kJ} \cdot \text{mol}^{-1}$ (PBE-D3+NPA), 1.9 $\text{kJ} \cdot \text{mol}^{-1}$ (PBE-D3+Chelpg), 4.6 $\text{kJ} \cdot \text{mol}^{-1}$ (TPSS-D3+NPA), and 1.3 $\text{kJ} \cdot \text{mol}^{-1}$ (TPSS-D3+Chelpg). The RMSDs are larger (12.7 , 19.3 , 12.7 , and 21.8 $\text{kJ} \cdot \text{mol}^{-1}$), but still notably smaller than those for CGFF. Nevertheless, the MM3 energies might significantly benefit from error compensation, namely the overestimation of cell volumes and as a consequence of intermolecular distances as well. Since a point charge description of electrostatic interactions is expected to perform better the farther charges are separated from the QM region, point charges very close to the central molecule are most likely to cause large errors. Hence, especially hydrogen bonded systems, having nearby charges at the donor/acceptor sites in the MM region, would show large errors in E_{coh} . It seems plausible that the overestimation of intermolecular distances counterbalances

Table 4.43: Cohesive energy deviation between ac(1)-QM/MM optimised structures at various levels of theory and the experimental results. All values in $\text{kJ} \cdot \text{mol}^{-1}$.

	PBE-D3/TZVP				TPSS-D3/TZVP			
	NPA		Chelpg		NPA		Chelpg	
	CGFF	MM3	CGFF	MM3	CGFF	MM3	CGFF	MM3
1,4-cyclohexanedione	-6.3	14.8	-9.2	11.5	-7.8	13.3	-10.5	10.1
acetic acid	-21.1	-5.4	-32.3	-19.7	-21.5	-6.3	-36.9	-21.5
adamantane	-1.8	20.4	-1.4	20.8	-2.0	20.3	-1.5	20.7
ammonia	-5.9	5.1	-24.9	-6.2	-3.9	6.4	-22.9	-5.2
anthracene	-15.2	-1.0	6.3	20.4	-12.3	1.8	6.7	20.8
benzene	-11.1	-2.4	0.3	8.0	-9.5	-1.0	0.6	8.3
carbon dioxide	-9.2	2.4	-11.3	0.8	-9.9	1.9	-11.6	0.6
cyanamide	-29.5	16.0	-89.4	-12.8	-27.1	13.1	-86.6	-14.9
cytosine	-62.8	27.8	-86.6	23.6	-62.5	27.6	-86.0	22.9
ethylcarbamate	-16.3	15.9	-19.9	15.4	-16.3	15.6	-19.8	14.8
formamide	-68.2	7.4	-70.0	10.2	-68.7	7.0	-73.2	10.0
hexamine	-4.2	12.3	-3.6	12.8	-3.0	13.1	-4.0	12.4
imidazole	-20.9	7.9	-24.1	8.0	-20.3	7.3	-24.3	8.3
naphthalene	-11.1	-0.1	3.1	14.4	-9.0	1.9	3.5	14.9
oxalic acid (α)	-35.7	-14.6	-64.5	-40.5	-38.3	-15.3	-67.1	-57.3
oxalic acid (β)	-43.1	-18.6	-55.7	-47.5	-44.9	-19.0	-57.1	-50.1
pyrazine	-6.7	5.2	6.9	16.7	-5.5	6.1	7.0	16.6
pyrazole	-7.1	4.3	-7.6	7.1	-5.7	5.6	-7.1	7.3
succinic acid	-61.5	-21.4	-68.7	-29.3	-62.5	-23.1	-70.7	-32.3
triazine	14.9	14.8	16.3	16.0	15.0	14.9	15.9	15.4
trioxane	-20.1	0.7	-4.2	14.6	-21.7	-0.6	-6.8	12.7
uracil	-77.4	11.6	-73.5	3.1	-79.6	11.3	-74.1	18.7
urea	-90.0	3.8	-111.5	-3.8	-89.3	3.9	-111.0	-3.9
MD	-26.5	4.7	-31.5	1.9	-26.4	4.6	-32.1	1.3
RMSD	37.9	12.7	48.2	19.3	38.1	12.7	48.6	21.8

the aforementioned error and is, thus, responsible for seemingly better cohesive energies with MM3. In the case of cyanamide, as a typical example of electrostatics dominating the intermolecular interactions, the closest intermolecular hydrogen bond (2.307 Å) is significantly longer in the PBE-D3/TZVP+NPA+MM3 optimised structure than in the respective CGFF result (1.903 Å). The respective QM energy in the point charge field is shifted upwards by $28.4 \text{ kJ} \cdot \text{mol}^{-1}$ for the MM3 structure with respect to CGFF, which is a significant part of the total $45.5 \text{ kJ} \cdot \text{mol}^{-1}$ shift. This already indicates that the change of electrostatic interactions between the QM molecule and its environment strongly contributes to the differences between MM3 and CGFF results. An evaluation of the deviation to experiment including only systems without hydrogen bonds provides further proof for this assumption. While the RMSD values for MM3 results slightly decrease (to 10.2, 14.7, 10.2, and $14.4 \text{ kJ} \cdot \text{mol}^{-1}$), the CGFF errors are dramatically reduced (to 11.4, 7.8, 11.0, and $8.2 \text{ kJ} \cdot \text{mol}^{-1}$). That the structures under consideration do not have point charges in close proximity to the QM molecule and their RMSDs are quite similar for CGFF and MM3 is considered as a strong indicator for the presence of intermolecular hydrogen bonds to be the main error source.

Albeit it seems somewhat favourable to use MM3 for further calculations in order to gain better energies, it is not done for one reason. The energies are more reasonable than the CGFF ones, but the structures are not. Once intermolecular interactions become very strong, as in the case of hydrogen bonded systems, the simple point charge representation of surrounding molecules becomes less reliable anyway, hence, an improved description is required. If that description is based on an ac(1)-QM/MM geometries, it seems more efficient to have rather reasonable geometries and calculate better energies on top, than to have unreliable geometries with accidentally well behaved ac(1)-QM/MM energies. Thus, CGFF parameters are deemed the more promising choice and are used as the default for further calculations.

Table 4.44: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} to experimental results calculated with different basis sets. Optimisations with ac(2)-QM/MM at the PBE-D3/Basis+NPA+CGFF (Basis= TZVP, QZVPP) level of theory. The cohesive energy is also calculated with a QZVPP basis at the TZVP optimised geometry (QZVPP@TZVP). Volumes in \AA^3 , energies in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV		ΔE_{coh}		
	TZVP	QZVPP	TZVP	QZVPP@TZVP	QZVPP
1,4-cyclohexanedione	-3.1	-2.2	-16.3	-8.5	-8.6
acetic acid	-4.0	-2.8	-14.4	-8.9	-8.9
adamantane	-4.7	-4.1	-12.3	-10.6	-11.3
ammonia	-13.0	-13.0	-25.1	-15.3	-15.4
anthracene	-2.2	-2.4	-6.2	-4.0	-4.3
benzene	-5.0	-4.9	-12.5	-10.6	-11.0
carbon dioxide	-1.5	-0.4	-5.7	-3.5	-3.5
cyanamide	-4.1	-3.0	-26.6	-22.2	-22.3
cytosine	-1.8	-0.8	-7.6	1.1	0.9
ethylcarbamate	-4.5	-3.9	-15.6	-9.1	-9.5
formamide	-4.5	-2.7	-17.6	-11.1	-11.5
hexamine	-4.2	-4.0	-20.8	-12.6	-12.7
imidazole	-2.1	-1.7	-22.3	-18.5	-18.6
naphthalene	-2.1	-1.7	-9.3	-7.2	-7.1
oxalic acid (α)	-2.8	-1.9	-20.1	-13.0	-13.1
oxalic acid (β)	-3.3	-1.3	-18.5	-11.8	-12.2
pyrazine	-6.4	-5.6	-17.8	-13.7	-13.7
pyrazole	-1.3	-0.9	-17.7	-13.8	-13.9
succinic acid	-2.4	-1.7	-22.2	-14.8	-15.7
triazine	-6.9	-6.3	-12.9	-8.1	-8.1
trioxane	-0.2	0.5	-17.3	-9.0	-8.9
uracil	-3.5	-3.0	-21.5	-14.1	-14.2
urea	-2.4	-2.4	-31.4	-22.6	-23.2
MD	-3.7	-3.1	-17.0	-11.4	-11.6
MAD	3.7	3.1	17.0	11.5	11.7
RMSD	4.5	4.1	18.2	12.6	12.8

4.8.2 ac(2)-QM/MM

The influence of both the population analysis and LJ parameters is expected to be much smaller at the ac(2)-QM/MM level, as the strongest interactions are described at the QM level. Hence, these are not varied, but restricted to NPA and CGFF. On the other hand, the influence of the QM method and the basis set is on the intermolecular interactions and, thus, the cell geometry and cohesive energy should be much larger than at the ac(1)-QM/MM level. Just as for ac(1)-QM/MM both functionals PBE-D3 and TPSS-D3 were employed, both with the TZVP and QZVPP basis sets. Since interactions from a supramolecular approach are expected to suffer from the BSSE, which results in overbinding, the basis size has a much larger impact on the intermolecular structure than in the ac(1)-QM/MM case. Therefore, ac(2)-QM/MM single-point calculations at the PBE-D3/QZVPP+NPA+CGFF level of theory were performed on the ac(2)-QM/MM PBE-D3/TZVP+NPA+CGFF optimised structures, as well as full ac(2)-QM/MM PBE-D3/QZVPP+NPA+CGFF geometry optimisations.

Influence of the Basis Set

The comparison among these results (Table 4.44) shows a consistent trend in the cell volumes. With the only exception of anthracene, all cell volumes increase with the larger QZVPP basis set, the RMSD between TZVP and QZVPP volumes is 0.9%. The changes are quite uniform, yielding an MD of 0.7% close to the RMSD value. This indicates that a slight improvement for all systems is gained, but no severe errors for the smaller TZVP basis are expected. The same trend is observed in the cohesive energies. These are always by a few $\text{kJ} \cdot \text{mol}^{-1}$ less binding with QZVPP than with TZVP. Again, the MD of $5.4 \text{ kJ} \cdot \text{mol}^{-1}$ and RMSD of $6.0 \text{ kJ} \cdot \text{mol}^{-1}$ are rather close together, proving that no drastic deviations between the different basis sizes occur. Interestingly, this difference seems to be mainly attributed to the basis size, but not to structural

Table 4.45: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} to experimental results calculated with different basis sets. Optimisations with ac(2)-QM/MM at the TPSS-D3/Basis+NPA+CGFF (Basis=TZVP,QZVPP) level of theory. Volumes in \AA^3 , energies in $\text{kJ} \cdot \text{mol}^{-1}$.

	ΔV		ΔE_{coh}	
	TZVP	QZVPP	TZVP	QZVPP
1,4-cyclohexanedione	-0.9	-0.0	-9.0	-1.9
acetic acid	-0.3	0.9	-7.6	-3.4
adamantane	-2.3	-2.3	-4.9	-4.5
ammonia	-10.3	-9.9	-16.5	-8.5
anthracene	-3.0	-2.6	-6.9	-6.0
benzene	-3.5	-3.5	-8.8	-7.8
carbon dioxide	5.2	7.8	-2.0	-0.5
cyanamide	-2.0	-0.9	-18.8	-15.4
cytosine	-0.4	0.6	-0.9	5.8
ethylcarbamate	-2.4	-1.5	-8.5	-3.7
formamide	-1.2	1.6	-12.0	-7.0
hexamine	-4.0	-3.1	-17.0	-8.4
imidazole	-0.2	0.2	-16.8	-14.0
naphthalene	-1.9	-1.5	-8.2	-6.7
oxalic acid (α)	-0.7	0.4	-12.8	-6.4
oxalic acid (β)	-0.6	0.2	-11.1	-6.3
pyrazine	-4.7	-4.2	-13.4	-9.8
pyrazole	-0.2	0.2	-12.9	-10.0
succinic acid	0.9	1.3	-12.3	-8.2
triazine	-4.5	-3.8	-8.5	-4.7
trioxane	2.5	3.4	-11.6	-4.2
uracil	-2.9	-2.0	-16.1	-10.5
urea	-1.6	-1.2	-23.2	-16.1
MD	-1.7	-0.9	-11.3	-6.9
MAD	2.4	2.3	11.3	7.4
RMSD	3.3	3.3	12.4	8.4

changes arising from the optimisation with different basis sets. For single-point calculations with the QZVPP basis set on the TZVP optimised structures the MD and RMSD values of 5.7 and 6.2 $\text{kJ} \cdot \text{mol}^{-1}$ are only slightly larger than the values from QZVPP optimised structures. The energy differences arising from the structural differences between TZVP and QZVPP optimised structures are consistently below 1 $\text{kJ} \cdot \text{mol}^{-1}$. These are deemed to be negligible.

The same qualitative behaviour is obtained with TPSS-D3 instead of PBE-D3 (Table 4.45). Although in some cases (e.g. benzene) the change from TZVP to QZVPP has almost no effect, on average the volumes increase by 0.8%. Quite large differences of 2.4% (carbon dioxide) and 2.8% (formamide) are found, but no signs of heavily underestimated volumes with the smaller TZVP basis set. Cohesive energies are systematically less binding as well. Changes in the range from 0.4 $\text{kJ} \cdot \text{mol}^{-1}$ (adamantane) to 8.5 $\text{kJ} \cdot \text{mol}^{-1}$ (hexamine) are observed, yielding MD and RMSD values of 4.4 and 5.0 $\text{kJ} \cdot \text{mol}^{-1}$. Just as for PBE a rather uniform improvement in the cohesive energies is obtained, but no unacceptably large errors seem to arise with a TZVP basis set.

Comparison to Experimental Results

A comparison of both PBE-D3 and TPSS-D3 results to the experimental findings (Table 4.46) reveals a somewhat different behaviour of the functionals. For TPSS both too large and too small cell volumes are obtained, while for PBE systematically too small values (with the only exception trioxane) are obtained. Hence, the respective MDs of -3.1% (PBE) and -1.0% (TPSS) are quite different, but the RMSD values are rather similar (4.1% for PBE and 3.3% for TPSS). While the smaller RMSD indicates a better agreement with experiment for TPSS, the more systematic underestimation of volumes with PBE suggests a physically sound behaviour.

An interesting trend is observed in the cohesive energies. Both functionals mostly overbind the systems,

Table 4.46: Cell volume differences ΔV and cohesive energy differences ΔE_{coh} to experimental results from ac(2)-QM/MM calculations at the Func/QZVPP+NPA+CGFF (Func=PBE-D3,TPSS-D3) levels of theory. Volume differences in %, energy differences in $\text{kJ} \cdot \text{mol}^{-1}$.

	PBE-D3		TPSS-D3	
	ΔE_{coh}	ΔV	ΔE_{coh}	ΔV
1,4-cyclohexanedione	-8.6	-2.2	-1.9	0.0
acetic acid	-9.4	-3.5	-3.4	0.9
adamantane	-11.3	-4.1	-4.5	-2.3
ammonia	-15.4	-13.0	-8.5	-9.9
anthracene	-4.3	-2.4	-6.0	-2.6
benzene	-11.0	-4.9	-7.8	-3.5
carbon dioxide	-3.5	-0.4	-0.5	7.8
cyanamide	-22.3	-3.0	-15.4	-0.9
cytosine	0.9	-0.8	5.8	0.6
ethylcarbamate	-9.5	-3.9	-3.7	-1.5
formamide	-11.5	-2.7	-7.0	1.6
hexamine	-12.7	-4.0	-8.4	-3.1
imidazole	-18.6	-1.7	-22.3	-2.1
naphthalene	-7.1	-1.7	-6.7	-1.5
oxalic acid (α)	-13.1	-1.9	-6.4	0.4
oxalic acid (β)	-12.2	-1.3	-6.3	0.2
pyrazine	-13.7	-5.6	-9.8	-4.2
pyrazole	-13.9	-0.9	-10.0	0.2
succinic acid	-15.7	-1.7	-8.2	1.3
triazine	-8.1	-6.3	-4.7	-3.8
trioxane	-8.9	0.5	-4.2	3.4
uracil	-14.2	-3.0	-10.5	-2.0
urea	-23.2	-2.4	-16.1	-1.2
MD	-11.6	-3.1	-7.2	-1.0
MAD	11.7	3.1	7.7	2.4
RMSD	12.8	4.1	9.1	3.3

but generally the differences to experiment are by some $\text{kJ} \cdot \text{mol}^{-1}$ smaller with TPSS. This yields a smaller RMSD of $9.1 \text{ kJ} \cdot \text{mol}^{-1}$ (TPSS) compared to $12.8 \text{ kJ} \cdot \text{mol}^{-1}$ (PBE). However, for both very similar maximum deviations of $-23.2 \text{ kJ} \cdot \text{mol}^{-1}$ (urea with PBE) and $-22.3 \text{ kJ} \cdot \text{mol}^{-1}$ (imidazole with TPSS) are found. The overall smaller differences indicate a slightly better agreement with experiment for TPSS, but the qualitative behaviour of PBE and TPSS is very similar. On the other hand, the computational cost of TPSS calculations compared to PBE was generally found to be by 25 to 50% larger. As this remains a limiting factor, at least for the two-body calculations of large molecules with the QZVPP basis set, all further benchmarking studies relied on PBE rather than on TPSS.

5 ac-QM/MM Benchmarks and Applications

5.1 Performance of the ac-QM/MM Models

After the determination of the presumably most reliable model, both with respect to internal ac-QM/MM parameters and external theory parameters, it shall be evaluated how the ac-QM/MM models perform overall. Hence, all X23 structures were fully optimised using different ac-QM/MM schemes uniformly at the most reasonable level of theory, which had turned out to be PBE-D3/QZVPP+NPA+CGFF. Moreover, a coupled cluster based description was applied within the ac(2)-QM/MM scheme in order to obtain the presumably most accurate but still affordable description of intermolecular interactions. These ac-QM/MM results are considered the best accessible of the method (at the GGA-DFT or WFT level), but they need comparison to some kind of reference to assess the absolute performance.

Two choices are considered as a reference, experiment and plane-wave DFT calculations. The former is the obvious choice, as the agreement with experimental results is an indicator for the soundness and reliability of the method. However, the comparison to experimental results is hampered in two ways. Firstly, the experimental cell volume are measured at different finite temperatures, while the ac-QM/MM calculations represent a system at 0 K. This will, even if a model describes the state at 0 K exactly, introduce a difference between theory and experiment. Secondly, the cohesive energy is hardly accessible experimentally. Hence, the reference cohesive energies are back-corrected values calculated from experimental enthalpies. This may introduce an error depending on the quality of the correction scheme. On the other hand, plane-wave DFT calculations are a viable choice for comparison, as these provide results of the same DFT functionals as employed in the ac-QM/MM method. Therefore, one may gain information on the performance of ac-QM/MM compared to a well-established and widely used method, while intrinsic contributions due to the use of a specific functional are ruled out. Although such comparison is not perfectly valid, since different types of basis sets are employed for plane-wave and ac-QM/MM calculations, it is assumed that for the QZVPP basis errors (e.g. the BSSE) are sufficiently small.

5.1.1 ac(1)-QM/MM

The benchmark results at the ac(1)-QM/MM level show essentially the same behaviour as the previous results with a smaller TZVP basis set. Molecular geometries are well described at the ac(1)-QM/MM level (Figure 5.1). An overall RMSD between cartesian coordinates (without hydrogen) of 0.019 Å indicates only minor differences of the optimised to the experimental structure. The difference is significantly smaller than the respective value of 0.045 Å that had been obtained from gas phase optimisations at the same level of theory. This indicates a significant improvement through the use of the ac(1)-QM/MM method.

The comparison of ac(1)-QM/MM cell volumes (Figure 5.2) to experimental ones shows quite large deviations with an RMSD of 6.5%. For some hydrogen bonded systems (ammonia, formamide, and urea) even deviations of more than 10% are obtained. However, a systematic underestimation of cell volumes can still be found. Only two cases (pyrazole and triazine) yield too large volumes. In general, hydrogen bonded systems show larger differences (RMSD of 7.5% compared to experimental values, 5.0% for the remaining set), but there is no clear hint that one class of molecules is described particularly well. In comparison to the plane-wave (PW) results, quite similar differences are obtained as in comparison to the experimental ones.

A more distinct difference is observed in the cohesive energies. The ac(1)-QM/MM energies (Figure 5.2) yield an RMSD of 38.8 kJ · mol⁻¹ with respect to experiment, while huge individual deviations up to

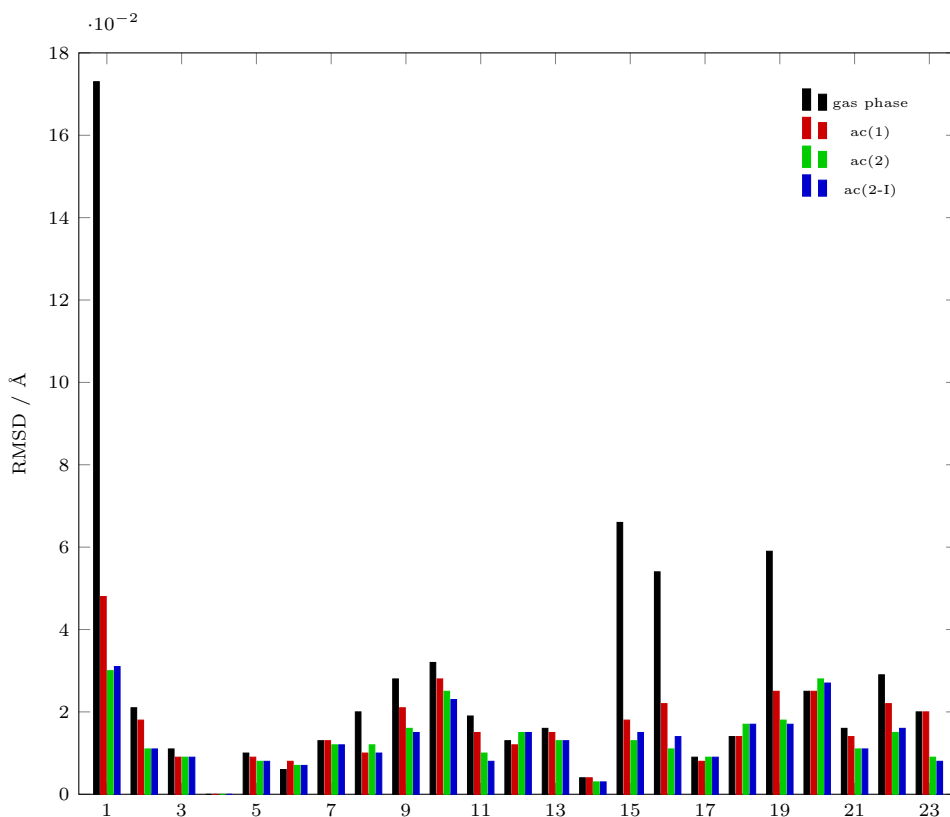


Figure 5.1: RMSD on cartesian atom coordinates (without hydrogen) of gas phase (gas phase), ac(1)-QM/MM (ac(1)), ac(2)-QM/MM (ac(2)), and ac(2-I)-QM/MM (ac(2-I)) results with respect to experimental values. All calculations at the PBE-D3/QZVPP+NPA+CGFF level of theory.

$-90.1 \text{ kJ} \cdot \text{mol}^{-1}$ (urea) are observed. In this case the RMSD for hydrogen bonded systems of $50.6 \text{ kJ} \cdot \text{mol}^{-1}$ is by a factor of four larger than the respective value $11.6 \text{ kJ} \cdot \text{mol}^{-1}$ for the residual systems. This indicates that even for polar molecules, like carbon dioxide, quite reasonable cohesive energies can be obtained at the ac(1)-QM/MM level. Hydrogen bonds, however, require an improved description of intermolecular interactions. A comparison to PW results shows slightly smaller overall differences, but no indication that ac(1)-QM/MM energies are much closer to PW results than to experimental ones.

Hence, it is clear that ac(1)-QM/MM can already provide reliable molecular geometries. Cell parameters and cohesive energies are also reasonable, unless intermolecular hydrogen bonds are present. In the latter case the cell geometries are still acceptable, but the energies are strongly overbinding.

5.1.2 ac(2)-QM/MM and ac(2-I)-QM/MM

DFT Results

The benchmark results with both two-body schemes ac(2)-QM/MM and ac(2-I)-QM/MM are distinctly different from the one-body results. Both variants yield essentially the same molecular geometries with the same RMSDs (without hydrogen atoms) of 0.015 \AA to experiment (Figure 5.1). This represents a small improvement over the respective one-body value of 0.019 \AA .

However, in the optimised cell volumes ac(2)-QM/MM and ac(2-I)-QM/MM results are distinctly different (Figure 5.2). Just as it had already been observed for the smaller TZVP basis, all ac(2)-QM/MM volumes are smaller than the experimental ones. Only in the case of trioxane a slightly larger value (0.5%) is obtained. With the QZVPP basis acceptable volume deviations of mostly below 4% are observed with an overall RMSD of 4.1%. It turns out that ammonia is still a difficult case with its large deviation of -13.0% . The residual systems without ammonia show a much smaller RMSD of 3.1%. A significantly larger deviation for ammonia than for the rest (6.8 and 1.4%, respectively) had been found in plane-wave calculations⁹⁸ using the same

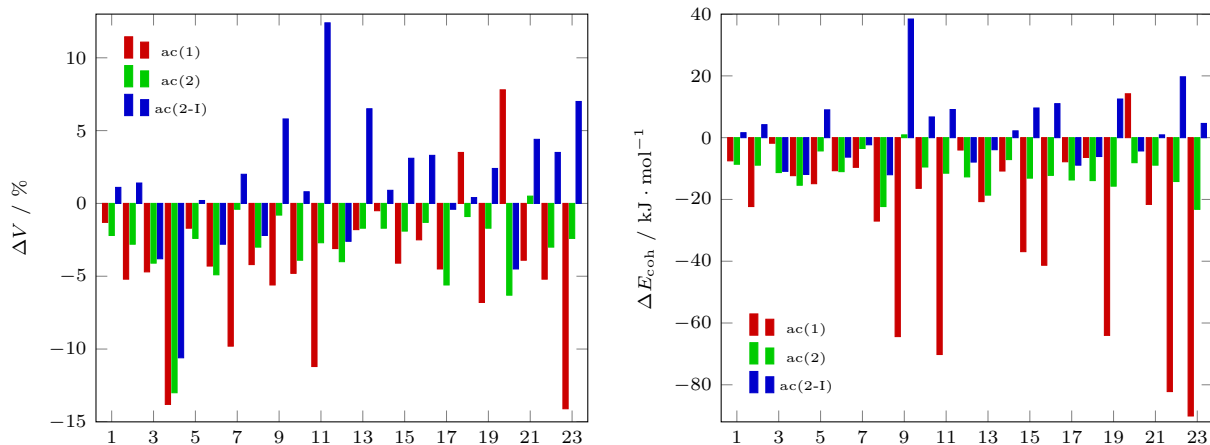


Figure 5.2: Cell volume deviations ΔV and cohesive energy deviations ΔE_{coh} of ac(1)-QM/MM (ac(1)), ac(2)-QM/MM (ac(2)), and ac(2-I)-QM/MM (ac(2-I)) optimisation results with respect to experimental values. All calculations at the PBE-D3/QZVPP+NPA+CGFF level of theory.

functional. Thus, ammonia is deemed a systematic outlier.

In contrast, ac(2-I)-QM/MM volumes behave significantly less uniform than ac(2)-QM/MM results. The majority of differences is positive, showing larger values up to 12.4% (formamide) compared to experiment. However, in some cases volumes are distinctly smaller for ac(2-I)-QM/MM. The RMSD to experiment of 4.7% is larger than for ac(2)-QM/MM, and especially the range of differences spans a quite large region from -10.6% (ammonia) to 12.4% (formamide), while for ac(2)-QM/MM deviations between -13.0% (ammonia) and 0.5% (trioxane) are obtained. The MD of 1.2% for ac(2-I)-QM/MM indicates that on average cell volumes are rather over- than underestimated. This is in contrast to the expectation of too small volumes.

A similar behaviour of the cohesive energy differences with respect to experiment is observed for both schemes. With the only exception of cytosine, for which an underbinding of $0.9 \text{ kJ} \cdot \text{mol}^{-1}$ is obtained, all energies show an overbinding in the ac(2)-QM/MM case. The deviations are much smaller than with ac(1)-QM/MM, but still go up to $-23.2 \text{ kJ} \cdot \text{mol}^{-1}$ for urea. In fact, the majority of differences lies in the range of -5 to $-15 \text{ kJ} \cdot \text{mol}^{-1}$, with an MD of $-11.6 \text{ kJ} \cdot \text{mol}^{-1}$ and an RMSD of $12.8 \text{ kJ} \cdot \text{mol}^{-1}$. In some cases the differences are very similar to ac(1)-QM/MM, but especially for some hydrogen bonded systems the ac(2)-QM/MM results are by several tens of $\text{kJ} \cdot \text{mol}^{-1}$ less binding and, therefore, in much better agreement with experiment.

For ac(2-I)-QM/MM, a very similar trend as for the volumes can be observed. A majority of cohesive energies are underbinding compared to the experimental results, but about one third is still overbinding. This averages to an MD of $2.4 \text{ kJ} \cdot \text{mol}^{-1}$. Nevertheless, the RMSD of $11.7 \text{ kJ} \cdot \text{mol}^{-1}$ is not much smaller than the value of $12.8 \text{ kJ} \cdot \text{mol}^{-1}$ obtained for ac(2)-QM/MM. Moreover, a much larger maximum deviation of $38.4 \text{ kJ} \cdot \text{mol}^{-1}$ (cytosine) is found compared to ac(2)-QM/MM ($-23.2 \text{ kJ} \cdot \text{mol}^{-1}$ for urea). A notably larger spread of the deviations of $50.4 \text{ kJ} \cdot \text{mol}^{-1}$ (-12.0 to $38.4 \text{ kJ} \cdot \text{mol}^{-1}$) for ac(2-I)-QM/MM than for ac(2)-QM/MM (spread of $24.1 \text{ kJ} \cdot \text{mol}^{-1}$, from -23.1 to $0.9 \text{ kJ} \cdot \text{mol}^{-1}$) is observed as well. Although there is no physically justified expectation of over- or underbinding, the qualitative agreement of ac(2)-QM/MM with ac(1)-QM/MM, including a systematic improvement in the energies at the ac(2)-QM/MM level, would rather support the use of ac(2)-QM/MM, not ac(2-I)-QM/MM.

DFT Results Revisited

Some additional information on the quality of the many-body approaches using PBE-D3 can be obtained from comparison to plane-wave (PW) results obtained with the same functional. The differences in volume and cohesive energy are plotted in Figure 5.3 and show some distinct features. For the cell volumes the ac(2)-QM/MM differences to PW are reasonably small (RMSD of 2.8%). Still ammonia, as the most challenging

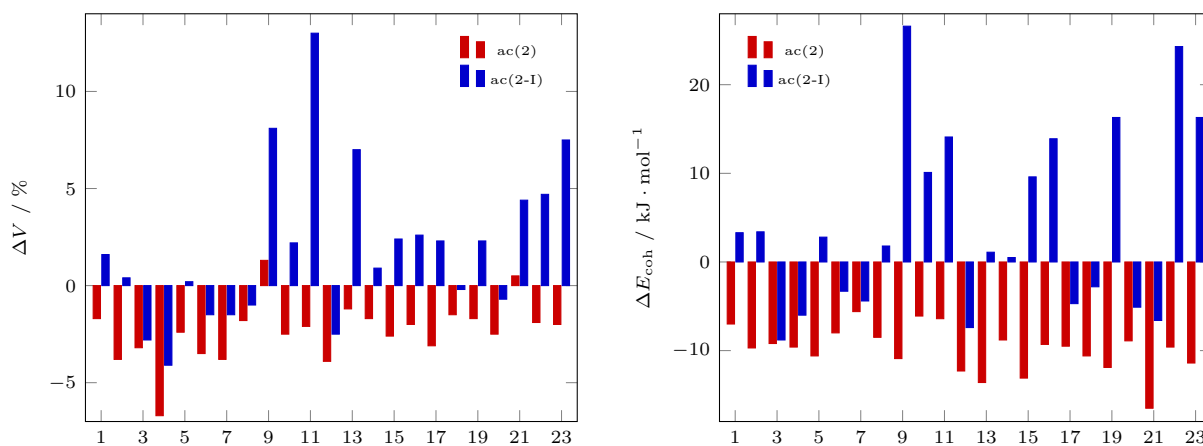


Figure 5.3: Cell volume deviations ΔV and cohesive energy deviations ΔE_{coh} of ac(2)-QM/MM (ac(2)) and ac(2-I)-QM/MM (ac(2-I)) results with respect to plane-wave PBE-D3 results. All ac-QM/MM calculations at the PBE-D3/QZVPP+NPA+CGFF level of theory.

system, is a notable outlier with a volume difference of -6.7% . For ac(2-I)-QM/MM, this behaviour cannot be observed at all. In that case a significantly larger RMSD of 4.4% to PW is obtained than for ac(2)-QM/MM. For ammonia a rather small deviation of -4.1% is obtained, but formamide with a large deviation of 13.0% is a notable outlier.

A more distinct effect can be observed in the cohesive energy differences to PW. For ac(2)-QM/MM an RMSD of $10.2 \text{ kJ} \cdot \text{mol}^{-1}$ is obtained. Interestingly, the differences become notably uniform. They span a range of only $10.9 \text{ kJ} \cdot \text{mol}^{-1}$, with largest and smallest differences of $16.5 \text{ kJ} \cdot \text{mol}^{-1}$ (trioxane) and $5.6 \text{ kJ} \cdot \text{mol}^{-1}$ (carbon dioxide). Consequently, the MAD and RMSD values are almost the same, namely 9.9 and $10.2 \text{ kJ} \cdot \text{mol}^{-1}$. This may be a consequence of small BSSE contributions from the different dimer calculations. Even sub- $\text{kJ} \cdot \text{mol}^{-1}$ errors per dimer might add up to several $\text{kJ} \cdot \text{mol}^{-1}$ for the total energy. Nonetheless, it is not clear if this is the main error source. In contrast, the ac(2-I)-QM/MM energy differences do not reveal a similar behaviour. The distribution of differences to PW is much broader, spanning a range of $35.4 \text{ kJ} \cdot \text{mol}^{-1}$. Deviations in both directions are found, from -8.8 up to $26.6 \text{ kJ} \cdot \text{mol}^{-1}$, yielding an MD of $4.1 \text{ kJ} \cdot \text{mol}^{-1}$. Nevertheless, the RMSD to PW ($10.9 \text{ kJ} \cdot \text{mol}^{-1}$) is significantly larger than the MD, and it is slightly larger than the ac(2)-QM/MM value.

The comparison of both approaches with PW results, therefore, leads to a distinct conclusions that the ac(2)-QM/MM variant is much more well-behaved, both in terms of volumes and energies. The uniform deviation in volumes and even more cohesive energies for ac(2)-QM/MM indicates a much more reliable, although systematic error. While it is not sure where this difference stems from, the assumption of added up BSSE contribution is in line with the systematic overbinding. However, no such systematic trend can be found for ac(2-I)-QM/MM at all. Therefore, the use of ac(2)-QM/MM seems more promising, especially in terms of consistency among different systems. It is used as the default choice for two-body calculations.

Coupled-Cluster Energies

While the ac-QM/MM model is designed to be inexpensive and flexible in order to describe medium to large sized molecular crystals, it can potentially be used to converge towards the many-body limit description of crystals with correlated wave function theory. Therefore, it is of interest to see how much can be gained from the use of such methods. Single-point LCCSD(T0)/VQZ+NPA+CGFF calculations were performed for the PBE-D3/QZVPP+NPA+CGFF optimised structures. NPA charges were calculated from the underlying HF calculation. However, for some systems the charges were also calculated at the PBE-D3/TZVP level of theory to obtain more comparable results to the ac(2)-QM/MM PBE-D3/QZVPP+NPA+CGFF energies. The results in Figure 5.4 unfortunately show no clear improvement of coupled cluster energies with HF based NPA charges (CC(NPA@HF)) or PBE based NPA charges (CC(NPA@PBE)). The RMSD to experiment decreases

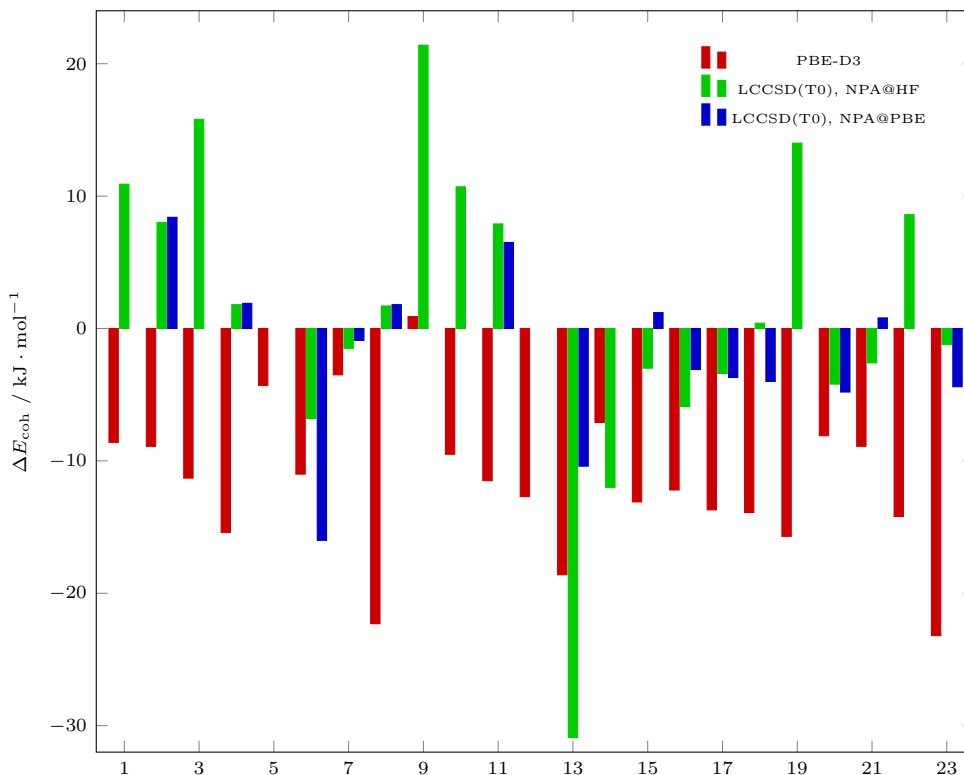


Figure 5.4: Cohesive energy deviations ΔE_{coh} of ac(2)-QM/MM calculations with respect to experimental results. The ac(2)-QM/MM energy calculations were performed at the PBE-D3/QZVPP+NPA+CGFF (PBE-D3) and LCCSD(T0)/VQZ+NPA+CGFF (LCCSD(T0)) levels of theory, both based on the ac(2)-QM/MM PBE-D3/QZVPP+NPA+CGFF optimised structures. The coupled cluster energies were obtained with NPA charges derived from either HF (NPA@HF) or PBE (NPA@PBE) calculations.

from $13.3 \text{ kJ} \cdot \text{mol}^{-1}$ (PBE) to $11.0 \text{ kJ} \cdot \text{mol}^{-1}$ (CC(NPA@HF)). If only the systems are taken into consideration for which both HF and PBE based NPA charges were used, the RMSDs are significantly reduced. For this subset a more distinct improvement from $14.2 \text{ kJ} \cdot \text{mol}^{-1}$ (PBE-D3/QZVPP+NPA+CGFF) to $9.3 \text{ kJ} \cdot \text{mol}^{-1}$ (CC(NPA@HF)) is observed, and a further reduction to $6.3 \text{ kJ} \cdot \text{mol}^{-1}$ (CC(NPA@PBE)) is achieved. There is a strong trend towards less binding energies with coupled cluster. The MD is $-11.0 \text{ kJ} \cdot \text{mol}^{-1}$ with PBE, but for CC(NPA@HF) a value of $2.1 \text{ kJ} \cdot \text{mol}^{-1}$ is obtained. Among the limited set, very similar MDs of $-2.8 \text{ kJ} \cdot \text{mol}^{-1}$ and $-1.9 \text{ kJ} \cdot \text{mol}^{-1}$ (CC(NPA@HF) and CC(NPA@PBE)) are obtained. For the majority of X23 molecules, CC cohesive energies (for both HF and PBE based NPA charges) are less binding than the PBE energies. This might indicate that the local approximation eliminates the BSSE influence and, therefore, yields energies closer to experiment.

The comparison of these results to the deviations of plane-wave PBE-D3 results (PW)⁹⁸ mostly provides the same behaviour. However, the MD of PW is smaller ($-1.8 \text{ kJ} \cdot \text{mol}^{-1}$) than the CC(NPA@HF) value ($2.1 \text{ kJ} \cdot \text{mol}^{-1}$), and the RMSD is much smaller for PW ($6.0 \text{ kJ} \cdot \text{mol}^{-1}$) compared to CC(NPA@HF) ($11.0 \text{ kJ} \cdot \text{mol}^{-1}$). The CC(NPA@HF) deviations might be deemed a broad scattering around the experimental results, which leads to the small mean deviation of $2.1 \text{ kJ} \cdot \text{mol}^{-1}$, but also to large differences up to $30.9 \text{ kJ} \cdot \text{mol}^{-1}$ (imidazole). The latter is greatly reduced for CC(NPA@PBE) (maximum deviation of $-16.0 \text{ kJ} \cdot \text{mol}^{-1}$ for benzene). This, as well as the aforementioned smaller energy RMSD values, indicates that PBE based charges provide an improved electrostatic embedding compared to HF based charges.

Although these CC(NPA@PBE) differences have not been obtained for the whole X23 set, the results indicate a sound behaviour. Still, differences up to $16 \text{ kJ} \cdot \text{mol}^{-1}$ are larger than what would be expected for this theory level. The RMSD for the limited set ($6.0 \text{ kJ} \cdot \text{mol}^{-1}$) is essentially the same as the RMSD of the PW calculations ($6.1 \text{ kJ} \cdot \text{mol}^{-1}$). Moreover, there is no clear trend among the deviations. Neither hydrogen bonded systems nor dispersion bound ones show larger differences, only a slight trend to larger deviations

Table 5.1: Comparison of experimental (Exp) cell volumes and cohesive energies with plane-wave PBE-D3 (PW) results and ac(2)-QM/MM results at the PBE-D3/QZVPP+NPA+CGFF (ac(2)[PBE]), and LCCSD(T)/VQZ+NPA+CGFF (ac(2)[CC]) level of theory. Volumes in \AA^3 , energies in $\text{kJ} \cdot \text{mol}^{-1}$.

	V				E_{coh}			
	Exp	PW	ac(2)[PBE]	ac(2)[CC]	Exp	PW	ac(2)[PBE]	ac(2)[CC]
ammonia	135.1	125.9	117.5	130.0	-37.3	-43.1	-52.7	-37.5
carbon dioxide	177.9	184.1	177.1	169.0	-27.2	-25.1	-30.7	-29.4
cyanamide	415.7	410.6	403.0	406.3	-80.0	-93.8	-102.3	-81.2
formamide	224.1	222.9	218.1	226.0	-79.1	-84.2	-90.6	-74.0

for larger molecules might be stated.

A couple of notes has to be made with respect to these results. One point to consider is the selection of the two-body region. The selection size had been determined with DFT and a smaller basis set. Therefore, it is not guaranteed that the CC energies are converged. It is not believed to have a large influence, but further studies are deemed advisable if one aims for precise energies.

A second point is the choice of structures. It is a common practice to evaluate energies at a high theory level for structures optimised at a low theory level. Nonetheless, the intermolecular distances are quite sensitive to the theory level. Hence, the CC interaction contributions might change significantly for structures optimised at a different theory level. Nonetheless, the results do not indicate a systematic deviation with respect to experiment. Therefore, this is not expected to be a major issue.

One concludes that within ac(2)-QM/MM the use of local coupled cluster energies based on DFT optimised structures provides a major improvement. If DFT based NPA charges are used for the electrostatic embedding, the energy differences to experimental results are greatly reduced. Nevertheless, quantitative agreement to experimental values can not be regularly achieved. Energy differences are slightly smaller than for plane-wave DFT, but not clearly superior.

Numerical Coupled Cluster Optimisations

It would be helpful to compare the coupled cluster energies of the DFT optimised structures to those obtained from structures actually optimised at the LCCSD(T0)/VQZ+NPA+CGFF level of theory. Unfortunately, analytical gradients were not available from the underlying coupled cluster calculations. Thus, the use of numerical ac(2)-QM/MM gradients for coupled cluster optimisations was enforced. Since this requires a full ac(2)-QM/MM energy calculation for every displacement, the applicability is limited to the smallest molecules, namely ammonia, carbon dioxide, cyanamide, and formamide. Nevertheless, for these systems fully optimised LCCSD(T0)/VQZ+NPA+CGFF geometries could be obtained, providing the results shown in Table 5.1. The cell volume for ammonia, which is particularly far off from the experimental value (135.1 \AA^3) at the DFT level (117.5 \AA^3), is much closer for the coupled cluster optimisation (130.0 \AA^3). No such improvement is found for the other systems, in the case of formamide even a slightly too large volume (0.8%) is obtained. However, the differences are calculated to experimental values at finite temperatures, which includes a contribution from thermal expansion. The deviations are reasonably small for both DFT and coupled cluster. Therefore, it is not really clear whether the closer volume is indeed closer to the unknown zero-point value.

Nevertheless, a much larger improvement is found in the cohesive energies. These are in excellent agreement with the experimental values, with a maximum deviation of $5.1 \text{ kJ} \cdot \text{mol}^{-1}$ for formamide. The other values are within chemical accuracy, yielding an RMSD of only $2.8 \text{ kJ} \cdot \text{mol}^{-1}$. Thus, they are significantly closer than those from plane-wave PBE-D3 calculations, for which an RMSD of $8.0 \text{ kJ} \cdot \text{mol}^{-1}$ is obtained. This demonstrates the potential accuracy possible with the ac(2)-QM/MM approach.

5.2 Applications of the ac-QM/MM Model

5.2.1 An Unusual Piano-Stool Complex

One interesting test case for the ac-QM/MM model is the piano-stool complex **A**, depicted in Figure 5.5. The bending of cyclopentadienide (Cp) hydrogen atoms in alkali metal complexes has been investigated theoretically already several decades ago.¹⁰⁴ Hydrogen atoms in such compounds are found to be bent away from the cation. This effect has been mainly attributed to electrostatic interactions.¹⁰⁵ For gas phase calculations the bending was observed in an earlier study on this system at the LMP2/VTZ level of theory,¹⁰ and could also be reproduced at the PBE-D3/TZVP level of theory in this work. However, for this structure high-resolution X-ray diffraction data allowed for the precise determination of hydrogen atoms bending towards the cation.¹⁰

The complementary theoretical study on the system¹⁰ had revealed that the inclusion of electrostatic interactions with the surrounding molecules strongly influences the bending angle. Therefore, it is likely that the experimentally determined inclination is a result of the intermolecular interactions. Consequently, this system is a well-suited test case for the ac-QM/MM model. One can compare whether the general structure is closer to experiment (in terms of coordinates' RMSD) than the gas phase results, but more importantly one can observe if the hydrogen bending can be reproduced. The latter is a distinct feature of the crystal structure and should be reproduced by a model for molecular crystals.

Results

A gas phase geometry optimisation of a single molecule at the PBE-D3/TZVP level of theory leads to a bending of the Cp hydrogen atoms away from the cations (by 5.0° on average, see Table 5.2), which confirms the previous findings. Significantly different results are obtained once the environment is included within the ac-QM/MM model. Already at the ac(1)-QM/MM level an inversion of the inclination angles can be observed for all hydrogen atoms. An average bending angle of 2.3° towards the cation side is measured, already quite close to the average of 1.6° from the experimental structure.

Further improvement is introduced by the addition of two-body corrections at the ac(2)-QM/MM level. The bending is still directed towards the cation, but the average bending angle decreases to 0.8°. The deviation in the average angle is slightly larger for ac(2)-QM/MM (0.8°) compared to ac(1)-QM/MM (0.7°). Nevertheless, the individual angles (see Table 5.2) are notably closer to experiment, with an RMSD of 2.0° for ac(1)-QM/MM and only 1.3° for ac(2)-QM/MM. It is, however, hard to estimate the error bars for experiment

Therefore, it can be concluded that the ac-QM/MM model already at the one-body level qualitatively reproduces the structure of this system. With a more accurate QM description of intermolecular interactions

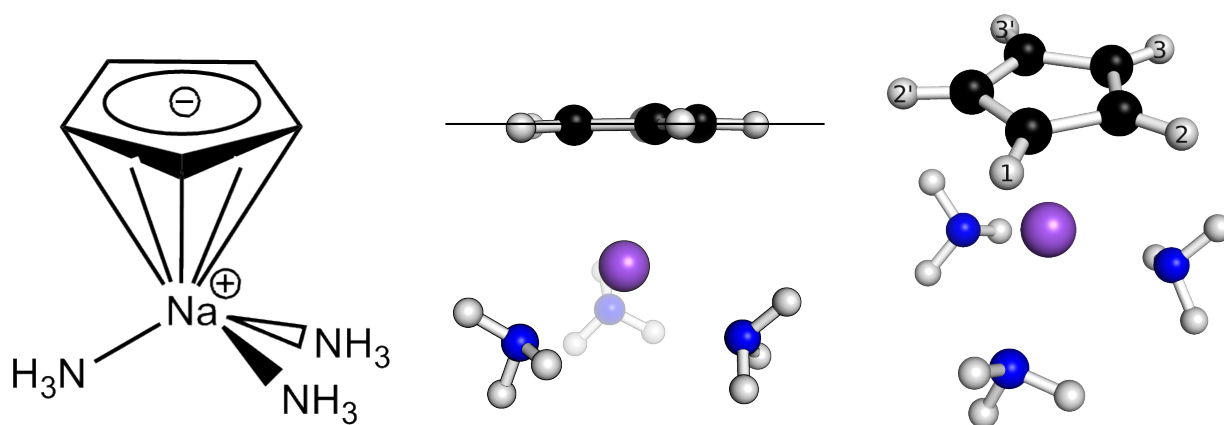


Figure 5.5: Lewis structure of the piano-stool complex **A**(left). The experimental structure shows a slight bending of the Cp hydrogen atoms towards the cation (center). Hydrogen atoms 2 and 2' are symmetry equivalent, as well as 3 and 3' (right).

Table 5.2: Comparison of out-of-plane bending of Cp hydrogen atoms in **A**. Comparison of experimental (Exp.) results^{96,97} with optimised structures for gas phase (GP), ac(1)-QM/MM (ac(1)), and ac(2)-QM/MM (ac(2)) results at the PBE-D3/def2-TZVP+NPA+CGFF level of theory. Values averaged over all hydrogen atoms (Average) and RMSDs between optimised and experimental angles (RMSD) are also shown. Angles and angle RMSDs are given in $^{\circ}$, positive angle values indicate bending towards the cation.

	H ₁	H ₂ ,H _{2'}	H ₃ ,H _{3'}	Average	RMSD
Exp.	2.2	0.4	2.6	1.6	–
GP	–6.0	–3.9	–5.6	–5.0	6.9
ac(1)	4.8	2.3	1.1	2.3	2.0
ac(2)	1.5	0.3	1.0	0.8	1.3

at the two-body level one can even achieve an almost quantitative agreement with the experimental results.

5.2.2 Investigation of a Diselenide X-ray Structure

State of Experiment

In the group of Prof. Dietmar Stalke the diselenide compound **B** (Figure 5.6) was investigated by X-ray diffraction. The refinement of their data showed an unexpected amount of residual density close to the selenium atoms. This density was uniformly observed for different crystals, experimental setups, and refinement models. Therefore, it could not be explained from an experimental point of view. Hence, a combined experimental and theoretical study was carried out.¹⁰⁶

The aforementioned residual density (see Figure 5.7) is well localised. Thus, a superposition of the main structure with a small amount of a disordered structure was assumed. There are four positions with such charge accumulations, each two of them are symmetry equivalent, which are referred to as charge peaks in the following. Since the peaks are spatially close to the selenium atoms, and selenium provides the largest amount of electron density of the elements in **B**, the peaks were considered to most likely be caused by disordered selenium atoms. However, the respective distances between atoms and charge peaks mostly do not match known selenium-selenium and selenium-carbon distances at all. Thus, a simple disorder of the unchanged molecule with similar atom distances was deemed unlikely.

An excited electronic state with a different geometry, especially with broken bonds, could be a possible reason for the finding. Since selenium is an X-ray fluorescent, it is possible that after relaxation an excited electronic state results. If this has a different geometry and a sufficient lifetime, it might evoke the charge peaks. Moreover, the compound was found to be active in electron paramagnetic resonance (EPR) spectroscopy after UV or X-ray irradiation on an hour timescale, thus, indicating the generation of a rather long-lived radical species.

Hence, a structure with selenium atoms at the charge peaks was considered as a possible disordered structure. In that case the Se–Se distance of 2.170 Å and one Se–C distance of 2.167 Å are still reasonable values for regular covalent bonds, while the second Se–C distance of 3.155 Å might indicate a broken bond, yielding a biradical structure (Figure 5.8). Albeit it is not clear how sound this guess is, since it includes no

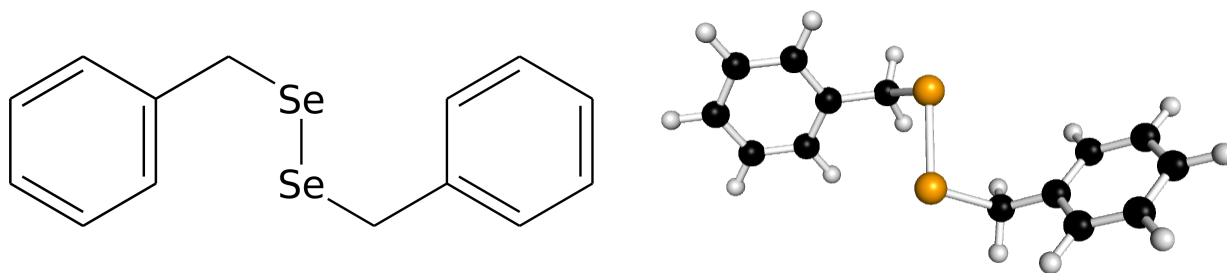


Figure 5.6: Lewis structure and X-ray structure of dibenzyl diselenide **B**.

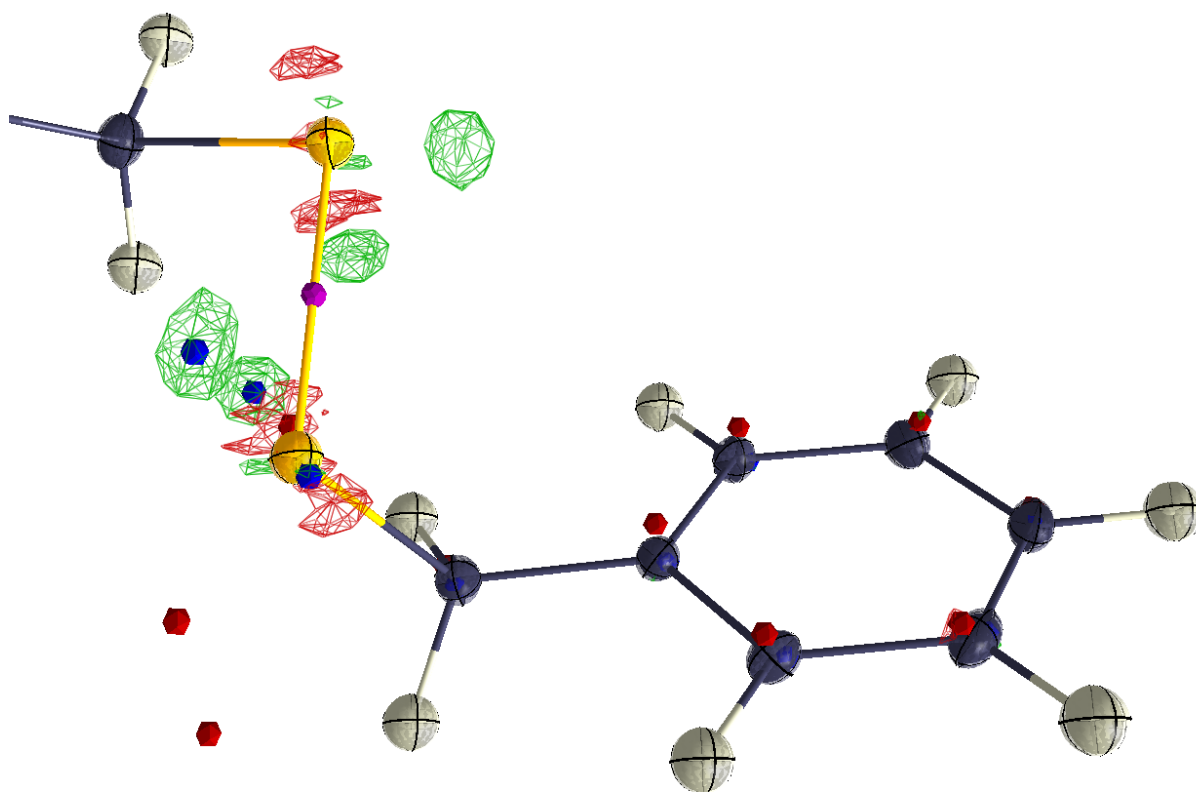


Figure 5.7: Residual density plot for the XD refinement of dataset D.¹⁰⁶ Isolevels $+0.3 \text{ e} \cdot \text{Å}^{-3}$ (green) and $-0.3 \text{ e} \cdot \text{Å}^{-3}$ (red). Density maxima (blue) and minima (red). The figure was prepared by and is presented with friendly permission of Christian Schürmann.

information on the carbon positions, it may still serve as a coarse approximation to validate optimisation results. It is referred to as the experimental guess in the following.

Adjustment of the ac-QM/MM Model

As a disordered molecule breaks the translational symmetry which is used in the ac-QM/MM model, some parts of the model need to be adjusted. It is assumed that only one molecule is disordered, hence, all other molecules are approximately the same as in a perfectly ordered crystal. Consequently, two optimisations are employed. Firstly, a standard ac(1)-QM/MM optimisation is performed, resulting in a relaxed structure of the ordered crystal. Then the environment is frozen (no changes to atomic coordinates and charges) and only the central molecule molecule is optimised in this static environment. This may now be calculated in an arbitrary electronic state for an arbitrary molecular charge. Consequently, the ac-QM/MM cohesive energy

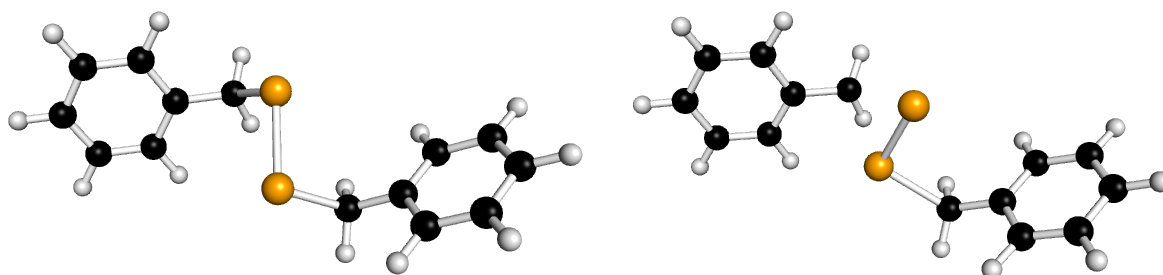


Figure 5.8: The main experimental structure (left) and the experimental guess (right) with selenium atoms moved to the charge peaks.

Table 5.3: Comparison of experimental (Exp.) geometry parameters with the optimised values calculated with ac(1)-QM/MM (ac(1)) and ac(2)-QM/MM (ac(2)) at the B3LYP-D3/TZVP+NPA+CGFF level of theory. Distances r in Å, angles \angle and dihedral angles dih in $^\circ$.

	$r(\text{SeSe}')$	$r(\text{SeC})$	$r(\text{Se}'\text{C}')$	$\angle(\text{Se}'\text{SeC})$	$\angle(\text{SeSe}'\text{C}')$	$dih(\text{CSeSe}'\text{C}')$
Exp.	2.315	1.984	1.984	101.1	101.1	93.6
ac(1)	2.328	1.996	1.996	101.3	101.3	95.6
ac(2)	2.335	1.995	1.995	101.4	101.5	93.6

expression is not applicable, as the central molecule is not the same as all other molecules. Unless one is explicitly interested in a cohesive energy of the disordered structure, it is possible to make use of relative energies. Under the restriction to a frozen environment only two energy contributions change compared to the ac-QM/MM energy. These are the internal energy of the central molecule and its interaction with the environment. They are readily included in the definition of E_{emb} (Equation 4.3). Therefore, a geometry optimisation with respect to E_{emb} should yield the energetically most favourable disordered configuration. The respective calculations are denoted as ac(1-D)-QM/MM.

Calculation Setups

In order to obtain a reasonable ordered structure, the charge peaks were not considered and the experimental geometry was optimised according to the ac(1)-QM/MM and ac(2)-QM/MM schemes at the B3LYP-D3/TZVP+NPA+CGFF level of theory. The ac(1)-QM/MM and ac(2)-QM/MM geometries (Table 5.3) are essentially the same and both very close to the experimental result. Moreover, the system **B** does not provide hydrogen bonds or very polar groups. Thus, it was considered reasonable to restrict all further calculations to the one-body level. The ac(1)-QM/MM optimised structure was used to define the frozen environment for all further embedded calculations.

Three sets of ac(1-D)-QM/MM optimisations were performed on the compound **B**. For the first set (Opt) the charge peaks were neglected and the ac(1)-QM/MM optimised structure was ac(1-D)-QM/MM optimised in the desired electronic state. The second set (Disp1) of calculations was also based on the ac(1)-QM/MM optimised geometry, but in this case the Se–Se distance of the central molecule was elongated prior to the ac(1-D)-QM/MM optimisation. This is expected to correspond to a biradical structure **B1** (Figure 5.9). For a third set (Disp2) of calculations the experimental guess was constructed from the ac(1)-QM/MM optimised structure (the selenium atoms moved to the positions of charge peaks). This corresponds to a biradical structure **B2** (Figure 5.9)

All ac(1-D)-QM/MM calculations were performed at the B3LYP-D3/TZVP level of theory with different electronic states taken into consideration. The singlet electronic ground state (S_0), the first excited singlet state (S_1 , via TD-DFT), lowest triplet state (T_0 , enforcing triplet multiplicity), and lowest broken-symmetry state (BrokenSym) were considered for the neutral molecules. Additionally, the structures were calculated as a cation or anion (always singly charged) as a doublet state (D_0).

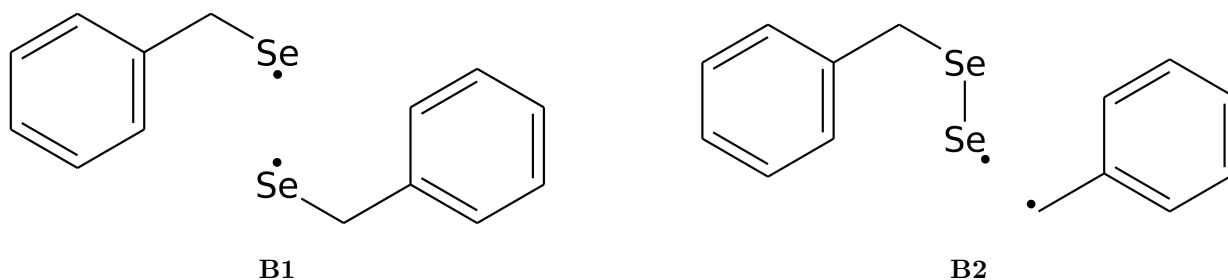


Figure 5.9: Lewis representation of two possible biradical structures arising from Se–Se bond dissociation (**B1**) or Se–C bond dissociation (**B2**).

Table 5.4: Comparison of geometry parameters between the ac(1)-QM/MM optimised structure (Ref.), the experimental guess based on the charge peaks (Guess) and structures from ac(1-D)-QM/MM optimisations of **B** at the B3LYP-D3/TZVP+NPA+CGFF level of theory in different electronic states. The S_0 , S_1 , T_0 , and broken symmetry (BrokenSym) states were calculated for the neutral molecule, the D_0 state for the singly charged molecule cation and anion. Distances r in Å, angles \angle and dihedral angles dih in $^\circ$.

	$r(\text{SeSe}')$	$r(\text{SeC})$	$r(\text{Se}'\text{C}')$	$\angle(\text{CSeSe}')$	$\angle(\text{SeSe}'\text{C}')$	$dih(\text{CSeSe}'\text{C}')$
Ref.	2.328	1.996	1.996	101.3	101.3	95.6
Guess	2.170	2.167	3.155	97.3	52.7	126.0
S_0	2.328	1.996	1.996	101.3	101.3	95.6
S_1	2.785	1.986	1.985	92.9	93.0	112.2
T_0	2.766	1.991	1.991	91.8	91.8	114.2
BrokenSym	2.328	1.998	1.998	101.6	101.6	95.1
D_0 (Cation)	2.268	2.056	2.056	101.1	101.0	132.1
D_0 (Anion)	2.862	1.983	1.984	95.9	95.3	85.3

Results

The results for the ac(1-D)-QM/MM optimisations of the unchanged ac(1)-QM/MM structures in Table 5.4 show no structural features close to those expected from the experimental guess. The ground state singlet optimisation precisely reproduces the ac(1)-QM/MM geometry. The broken symmetry calculation always indicates pure singlet character and, hence, results in the same structure, with only tiny differences of less than 0.002 Å in the distances and less than 0.5° in the angles. The triplet and excited singlet states, however, both yield different structures albeit very similar. In both cases the main difference to the ground state result is a significantly elongated Se–Se distance of 2.785 Å (T_0) and 2.766 Å (S_1) compared to 2.328 Å (S_0). The Se–C distances are different by only about 0.01 Å compared to S_0 . Moreover, both Se–C distances are equal, both for T_0 and S_1 . These findings do not agree with the experimental guess. The same is observed with the angles. Only the C–Se–Se–C dihedral angles of 112.2 (S_1) and 114.2 (T_0) indicate a geometry relaxation towards the experimental guess. The cation calculation does provide quite similar results as the ground state, only the C–Se–Se–C dihedral angle strongly increases from 95.6° to 132.1°. Much larger differences are found in the anionic case. The Se–Se distance increases from 2.328 to 2.862 Å, but the Se–C distances change by only 0.013 Å. The Se–Se–C angles and C–Se–Se–C dihedral angle decrease notably by up to 10°, but none of the changes is in line with the expected changes from the reference towards the experimental guess. From these optimisations there is no indication that any of the electronic states leads to a structure similar to the experimental guess. However, the calculations all start from the ground state equilibrium structure, but different results may be obtained from an already disordered starting structure.

If the selenium atoms are moved apart before the ac(1-D)-QM/MM optimisation is performed, the results (Table 5.5) are essentially the same as for the optimisation without the selenium displacement. The differences between the analogous results of both optimisations are mostly well below 0.002 Å for distances and 1.0° for angles, which are considered negligible. Only for the first excited singlet state the geometry parameters are significantly different. The C–Se–Se angles of 100.2 and 102.2° are larger compared to 92.9 and 93.0° in the Opt case, and the C–Se–Se–C dihedral angle of 73.4° is much smaller than the 112.2° in the Opt case. The differences in the monitored distances are less than 0.03 Å, but it is still likely that these are not the same structures. Nevertheless, none of these shows a significant breaking of the molecular symmetry with distinct changes of the atomic distances that could explain the assumed disordered structure.

However, the situation significantly changes if the ac(1)-QM/MM optimised starting structure is modified by shifting the selenium atoms to the charge peak positions. The optimisation as a singlet state converges smoothly to the reference structure, indicating that there is no barrier which artificially locks the molecule in the displaced geometry. Optimisation as a broken symmetry state starts at a triplet state, but converges to the singlet state at the reference geometry as well. However, once an electronic triplet state is enforced, the optimisation converges to a distinctly different energy minimum. In that case one of the Se–C distances (3.132 Å) remains much larger than the other (2.004 Å), and the respective Se–Se–C angle is much smaller (79.7°) than in the reference structure. The Se–Se distance (2.243 Å) is closer to the experimental guess

Table 5.5: Comparison of geometry parameters between the ac(1)-QM/MM optimised structure (Ref.), the experimental guess based on the charge peaks (Guess) and structures from ac(1-D)-QM/MM optimisations of **B1** at the B3LYP-D3/TZVP+NPA+CGFF level of theory in different electronic states. The S_0 , S_1 , T_0 , and broken symmetry (BrokenSym) states were calculated for the neutral molecule, the D_0 state for the singly charged molecule cation and anion. Distances r in Å, angles \angle and dihedral angles dih in $^\circ$.

	$r(\text{SeSe}')$	$r(\text{SeC})$	$r(\text{Se}'\text{C}')$	$\angle(\text{CSeSe}')$	$\angle(\text{SeSe}'\text{C}')$	$dih(\text{CSeSe}'\text{C}')$
Ref.	2.328	1.996	1.996	101.3	101.3	95.6
Guess	2.170	2.167	3.155	97.3	52.7	126.0
S_0	2.329	1.998	1.998	101.6	101.6	95.6
S_1	2.819	1.979	1.977	100.2	102.2	73.4
T_0	2.766	1.991	1.991	92.4	92.3	114.8
BrokenSym	2.329	1.998	1.999	101.6	101.6	95.6
D_0 (Cation)	2.284	2.041	2.041	106.9	106.9	80.0
D_0 (Anion)	2.860	1.983	1.983	94.5	94.5	85.8

Table 5.6: Comparison of geometry parameters between the ac(1)-QM/MM optimised structure (Ref.), the experimental guess based on the charge peaks (Guess) and structures from ac(1-D)-QM/MM optimisations of **B2** at the B3LYP-D3/TZVP+NPA+CGFF level of theory in different electronic states. The S_0 , S_1 , T_0 , and broken symmetry (BrokenSym) states were calculated for the neutral molecule, the D_0 state for the singly charged molecule cation and anion. Distances r in Å, angles \angle and dihedral angles dih in $^\circ$.

	$r(\text{SeSe}')$	$r(\text{SeC})$	$r(\text{Se}'\text{C}')$	$\angle(\text{CSeSe}')$	$\angle(\text{SeSe}'\text{C}')$	$dih(\text{CSeSe}'\text{C}')$
Ref.	2.328	1.996	1.996	101.3	101.3	95.6
Guess	2.170	2.167	3.155	97.3	52.7	126.0
S_0	2.328	1.997	1.998	101.3	101.6	94.8
S_1	—	—	—	—	—	—
T_0	2.243	2.004	3.132	100.8	79.7	109.2
BrokenSym	2.328	1.998	1.998	101.5	101.6	94.7
D_0 (Cation)	2.284	2.040	2.040	106.5	106.4	79.4
D_0 (Anion)	2.325	1.994	2.734	99.9	95.1	98.4

(2.170 Å) than for all other ac(1-D)-QM/MM optimised structures. Moreover, the Se–C distances of 2.004 and 3.132 Å agree qualitatively with the experimental guess (2.167 and 3.155 Å). Similarly, for the C–Se–Se angles (100.8 and 79.7 $^\circ$) no quantitative agreement can be observed with the experimental guess (97.3 and 52.7 $^\circ$). Nonetheless, at least a qualitative agreement in several geometry parameters can be stated for the triplet state with the experimental guess. If an excited singlet state is targeted, the optimisation does not converge successfully, as it encounters a crossing with the singlet ground state. If the triplet result is chosen as a starting structure, the same outcome is observed. Therefore, it is assumed that there is no energy minimum for the first electronic state nearby. The optimisation runs with the molecule as a singly charged cation in a doublet state resulted in the very same structure which had been obtained in the Opt and Disp1 cases. No structural similarity to the experimental guess is found. For the respective optimisation as an anion, however, a somewhat qualitative agreement with the experimental guess is found. Although the Se–Se distance is very close to the ac(1)-QM/MM reference structure, two distinctly different Se–C distances are obtained. The shorter one is very close to all other ac(1-D)-QM/MM results, and the much longer distance of 2.734 Å is not very close to the experimental guess of 3.155 Å, but it is much closer than to the S_0 reference value (1.996 Å).

Considering all optimisations performed, only two calculations show a qualitatively correct trend for distances and angles, namely the triplet state and the anionic doublet state. In both cases for one Se–C distance very similar differences of 0.171 Å (triplet) and 0.173 Å (anion) to the guess are obtained. However, for the Se–Se distance, in the triplet case a much smaller difference of 0.073 Å is observed than in the anionic case with 0.155 Å. Similarly, the second Se–C distance difference is much smaller (0.023 Å) for the triplet optimisation than for the anionic calculation (0.421 Å). Consequently, from these results a disordered structure resulting from a triplet state is the best guess to explain the experimental charge peaks. Albeit the respective structure is not in perfect agreement with the experimental guess, the qualitative changes with respect to the singlet ground state indicate that this is a reasonable explanation.

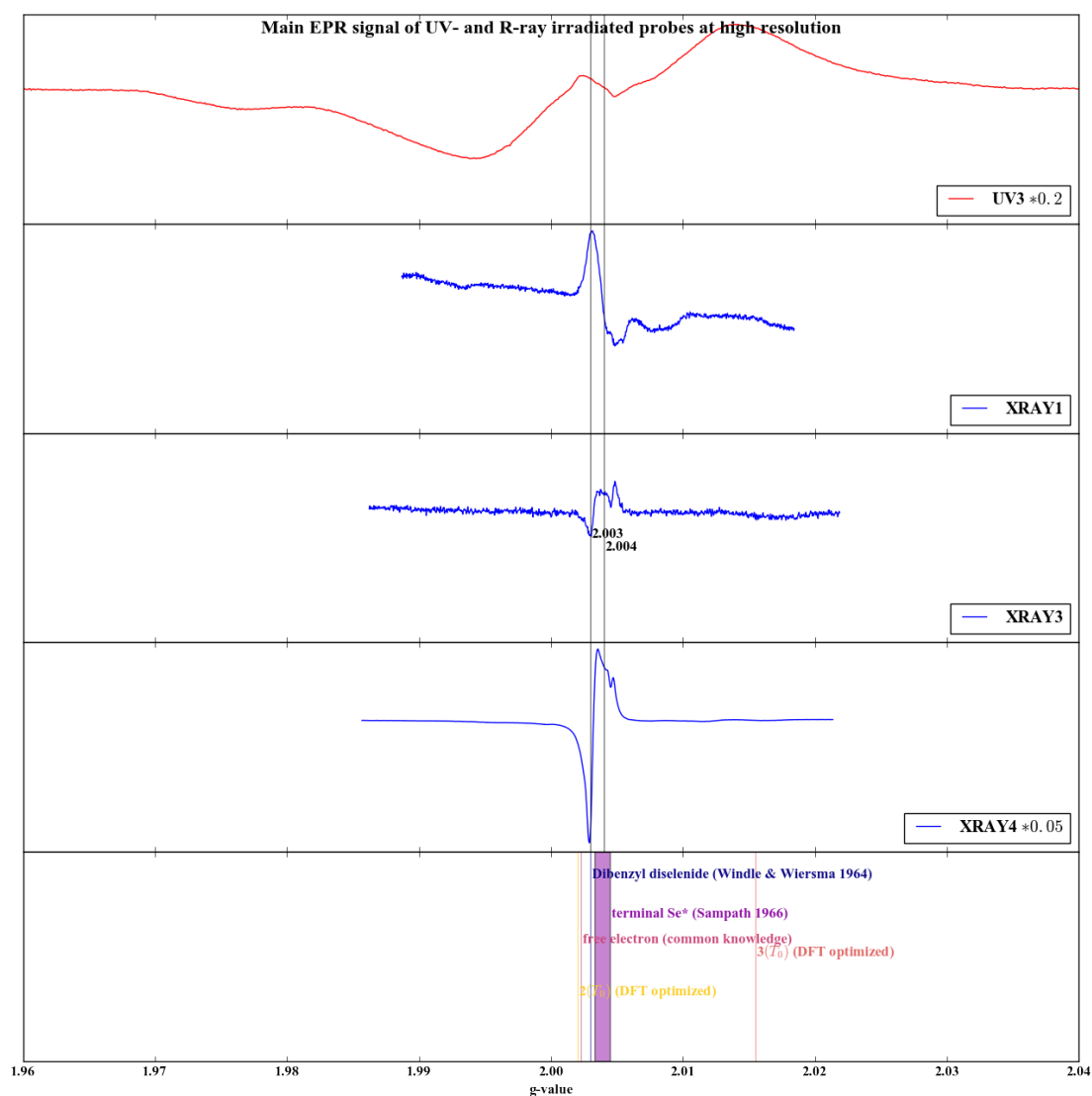


Figure 5.10: Experimental EPR spectra of **B**, focused at the main signal. Experimental spectra were recorded from molecular crystals of **B** under UV (UV3) or X-ray irradiation (XRAY1, XRAY3, XRAY4). Comparison is made to literature values of **B** (Dibenzyl diselenide),¹⁰⁷ terminal Se radicals (terminal Se*),¹⁰⁸ the free electron (free electron), and the theoretically determined g -values of **B2** (broken Se–C bond, $2(T_0)$) and **B1** (broken Se–Se bond, $3(T_0)$) in the T_0 state. The graphic was prepared by and is presented with friendly permission of Christian Schürmann.

In order to gain additional proof of a radicalic structure **B2**, (Figure 5.9), EPR parameters were calculated. The g -tensors were calculated at the B3LYP-D3/TZVP level of theory for the T_0 electronic state. The structures obtained from ac(1-D)-QM/MM optimisations of both **B2** and **B1** in the T_0 state were used unchanged.

The resulting EPR parameters and electronic structure data were compared to the experimental data collected and interpreted at the group of Prof. Dietmar Stalke. Calculated parameters of **B2** and **B1**, as well as known experimental parameters for UV irradiated **B**¹⁰⁷ and terminal Se radicals,¹⁰⁸ were compared to the experimental findings. Firstly, the main EPR signal at a g -value of 2.003 is in agreement with the signal obtained from UV irradiation of compound **B**.¹⁰⁷ The value is typical of aromatic radicals, e.g. benzyl radicals.¹⁰⁹ Essentially the same value is obtained from EPR calculations of the free benzyl radical as well. A second EPR signal is experimentally observed at 2.095. This is in agreement with experimental findings assigned to a Se radical comparable to **B1**.¹⁰⁷ The calculated g -factors for the Se radical in **B1** (2.073 and 2.158) are in the same region. Hence, it is not clear whether these signal stems from a Se radical of **B2** or **B1**. The latter might as well be present in the crystal, although no corresponding charge peaks were observed. Due to its structural similarity to **B** it is less likely to be found in the X-ray diffraction data. This

might, however, result for different reasons. Shorter lifetime of **B1**, a larger fraction of **B2** compared to **B1**, and smaller geometrical structural differences between **B** and **B1** could all make existing charge peaks less likely to be observed or recognised.

Although the EPR results are no solid proof for the existence of **B2**, they are in line with the expectation for the presence of **B2** in the crystal. Therefore, one is quite sure that the excitation of an electronic triplet state under X-ray irradiation takes place. The excited molecule relaxes to yield the metastable **B2** structure. The density of this structure, especially of its selenium atoms, is responsible for the charge peaks observed in the X-ray diffraction study of **B**.

6 Summary

An additive crystal QM/MM model (ac-QM/MM) for the calculation of molecular crystals has been developed and tested. While different QM/MM schemes for such purposes exist, this model provides three distinct advantages. Only a single molecule and its interactions with all other molecules need to be calculated. Most other schemes use a unit cell as the smallest unit. This allows for the reduction of computational cost by a factor equal to the number of molecules per cell in the ac-QM/MM model. Secondly, other models often only enable the optimisation of molecular structure or crystal structure, if at all. The ac-QM/MM model allows for an unrestricted optimisation of both. The third advantage is the flexibility. One can conveniently change the MM and QM model, the QM programme, and even combine different QM suites for different contributions. Other models are usually coded within an existing programme suite and are, therefore, restricted to the methods available within that program.

The ac-QM/MM model makes use of the fact that all equivalent molecules in a crystal behave in the same way. Therefore, only a single molecule is treated at the QM level, while a finite cluster of neighbouring molecules is treated at the MM level. These molecules are represented with point charges and Lennard-Jones parameters, additional point charges are fitted to represent the periodic electrostatics. Charges are QM derived, while MM parameters are taken from a force field. QM charges are determined in a self-consistent manner for each geometry, thus, updating the electrostatic potential after every optimisation step. The whole crystal is considered as multiple copies of the QM molecule, defined by the crystal symmetry. Therefore, only the QM molecule and its interactions with the environment are considered. The latter interactions may partly be replaced with a QM description to yield improved results.

Analytical expressions are derived to calculate ac-QM/MM gradients from a standard QM/MM gradient calculation. Gradient contributions for both QM and MM atoms are included in order to represent the update of all cluster atoms according to the crystal symmetry. Albeit symmetry is not explicitly enforced, symmetry relations between the QM molecule and all other molecules are used to project gradient contributions on the central molecule. The ac-QM/MM gradients allow for the unrestricted optimisation of the QM molecule, and the projection of these geometry changes determine the according update of all residual molecules.

Not only atomic gradients are available, but also stresses. These are analytically derived from the QM/MM gradients. They are used in the sense of gradients to allow for the optimisation of cell parameters. Although an ac-QM/MM internal optimisation of the cell is possible, the whole optimisation process is more efficiently performed by an external optimiser. Strain values are represented with the cartesian coordinates of two dummy atoms, while stresses are scaled and treated as the respective gradients. This allows the external optimiser to form an optimisation step that is applied internally in the ac-QM/MM scheme. Analytical gradients and stresses are found to agree very well with their numerical counterparts.

Nonetheless, a small contribution to the ac-QM/MM gradients and stresses can not be captured. The environmental charges are allowed to change between optimisation steps, but no corresponding contribution to the gradient is available. Hence, analytical gradients can only be successfully applied to the optimisation in conjunction with an adjusted charge update procedure. The update of the electrostatic potential of the environment charges is crucial and must not simply be neglected. The application of a tight threshold, especially without the application of a damping scheme, severely hampers the convergence. On the other hand, the neglect of the update leads to significant errors in geometry and energy. A loose threshold for the charge update in conjunction with a damping procedure is found to be a sound compromise. The damping depends on the energy difference between previous optimisation steps to ensure a robust behaviour at the end of a geometry optimisation.

A series of parameters was varied to establish a reasonable model. The X23 benchmark set was used to compare geometries and cohesive energies to experimental values. At the one-body level, the combination of NPA charges with CGFF parameters is deemed most reliable in terms of energies and geometries. Among the tested theory levels (PBE and TPSS functionals, TZVP and QZVPP basis sets), only minor differences in the results are found. Reasonable geometries and energies can be obtained at the PBE-D3/TZVP+NPA+CGFF level of theory for most systems. Nonetheless, hydrogen bonded systems impose a great challenge. Although the geometries are not unreasonably different from the experimental ones, the differences are much larger than for the residual systems. Cohesive energies even deviate by several tens of $\text{kJ} \cdot \text{mol}^{-1}$ from the experimental values and can, thus, not be considered reliable.

The picture changes upon the consideration of two-body QM corrections. TPSS-D3 is observed to yield slightly better energies and geometries than PBE-D3, but at a significantly increased computational cost. The influence of the basis set superposition error imposes systematically too small cell volumes and too strongly binding cohesive energies. Its impact is significantly reduced with a large QZVPP basis set, so, reasonable cell volumes and energies are obtained at the PBE-D3/QZVPP level. However, the energy differences to experiment are still quite large. A series of modifications to the scheme (freezing the wavefunction, manipulation of the environment charges) was tested to improve the energies. Beneficial error cancellation in some of these modifications could yield seemingly improved cohesive energies, but at the price of inferior geometries. Even for the most promising variants, a much less consistent behaviour of energies and volumes was observed. This includes unreasonable cell parameters and much larger spreads in the cohesive energies among the X23 set, thus, indicating an inferior reliability.

Benchmarking of the ac(1)-QM/MM scheme with the X23 set proves a reliable behaviour for all but the hydrogen bonded systems. Improved molecular geometries (compared to gas phase calculations) are available for all systems. Moderate differences to experiment (up to $22 \text{ kJ} \cdot \text{mol}^{-1}$) in the cohesive energies are obtained for the systems without hydrogen bonds. Much larger deviations (up to $90 \text{ kJ} \cdot \text{mol}^{-1}$) are found for hydrogen bonded systems. Moreover, in these cases large deviations in some cell parameters are observed.

For the ac(2)-QM/MM scheme, reliable and consistent results are obtained for all X23 systems with a DFT treatment (PBE-D3/QZVPP+NPA+CGFF). Molecular geometries are slightly improved compared to the ac(1)-QM/MM scheme, and cell parameters are much closer to experimental results. As expected from the neglect of thermal effects in the model, cell volumes are systematically smaller (typically by 2-4%) than experimental ones. Cohesive energies are systematically overbinding, up to $23 \text{ kJ} \cdot \text{mol}^{-1}$. In comparison to plane-wave PBE-D3 results, energy differences are found to be very similar across the X23 set. For all systems a deviation to the plane-wave energy in the range of $5\text{-}16 \text{ kJ} \cdot \text{mol}^{-1}$ is observed, highlighting a very consistent behaviour of the ac(2)-QM/MM model.

If energies are calculated at the coupled cluster level (LCCSD(T0)/VQZ+NPA+CGFF), significantly improved results are obtained. Calculations on the PBE optimised geometries could reduce the cohesive energy root mean square difference from 12.8 to $11.0 \text{ kJ} \cdot \text{mol}^{-1}$ if NPA charges of the HF wave-function are used. If NPA charges are based on a PBE calculation, the difference decreases further from $9.3 \text{ kJ} \cdot \text{mol}^{-1}$ to $6.3 \text{ kJ} \cdot \text{mol}^{-1}$ for a subset of the X23 molecules. Full numerical coupled cluster optimisations of the smallest systems could even provide highly accurate cohesive energies within $6 \text{ kJ} \cdot \text{mol}^{-1}$ to experiment. This accuracy is not routinely accessible, but proves that results can be systematically improved.

Application of the ac-QM/MM model to a sodium cyclopentadienide compound could reproduce its structural features. From X-ray diffraction, a bending of the cyclopentadienide hydrogen atoms towards the cation has been determined. This is in contradiction with both expectation from electrostatic repulsion and theoretical prediction from a gas phase model. The ac-QM/MM model, however, can qualitatively reproduce the experimental binding already at the one-body level. At the two-body level, the agreement with experiment is further improved.

In the case of dibenzylselenide the application of the ac(1)-QM/MM scheme could aid the interpretation of X-ray data. Experimental data revealed well-localised charge concentrations that could not be explained

for the ground state structure. Optimisations were carried out at the ac(1)-QM/MM level for the electronic ground state to establish a consistent environment. To investigate disordered structures, the environment was frozen and the QM molecule was re-optimised. From optimisations at different electronic states one structure reasonably fits to the experimental findings. It is a triplet state with a broken Se–C bond and, hence, significantly different Se atom positions. A small fraction of this disordered structure can reasonably explain the experimental charge concentration. Comparison of calculated and experimental EPR data further confirms the presence of this structure in the crystal.

Therefore, the ac-QM/MM scheme has proved to be a convenient and inexpensive tool for the calculation of molecular crystals. It can not only improve upon a simple gas-phase treatment at very moderate additional cost, but also allow for a systematic convergence to very accurate results upon the use of increasingly sophisticated QM methods.

7 Outlook

A series of improvements upon the existing ac-QM/MM model are possible. One point to be noted is the treatment of crystal symmetry. The current scheme includes implicit symmetry of the embedding potentials and explicit symmetry relations between the central molecule and the other cell molecules. This has proven to work well for all systems considered in this work. However, it requires some caution in the practical definition of molecules. This limits the ease of using the method, and can in conjunction with very symmetric structures lead to error. In the case of carbon dioxide, bent molecules were obtained once, as an improper atom numbering resulted in an incorrect composition of the gradients. This is most likely a very rare problem, as both the carbon dioxide molecule and its crystal represent a highly symmetric system. However, the explicit use of symmetry could help to prevent such cases. Moreover, it could make the method more convenient to use, as all required information could be read from a crystal information file (.cif).

Another point to be improved is the choice of LJ potentials. The CGFF parameters used in this work were found to yield reasonable structures and energies. Nevertheless, a system-specific determination of such parameters could be more appropriate. Several approaches for the determination of force field parameters from QM calculations exist. Also programmes for the automated determination of MM parameters on-the-fly from QM calculations are available.^{110,111,112} Unfortunately, these do not necessarily allow for a computationally feasible calculation. Force matching algorithms e.g. would require the calculation of a QM Hessian to fit the MM forces to. This could heavily increase the computational cost for each optimisation step. Moreover, these approaches usually aim to fit a set of force field parameters (bond stretching, angle bending, vdW, Coulomb, ...) which are mostly not needed for the ac-QM/MM model. Therefore, the vdW terms of such an approach need not perform better in the ac-QM/MM model than the standard values, even if the performance of the QM derived force field is superior to the standard force field. A fitting procedure for a vdW force fields¹¹³, using only vdW and Coulomb terms, might be appropriate for the ac-QM/MM model. Therefore, it seems most promising to calculate QM derived LJ parameters independently, and then use these parameters unchanged for a full ac-QM/MM optimisation. This could ideally allow for significantly improved cell parameters and cohesive energies at the one-body level.

A problem the ac-QM/MM has not been tested on so far are co-crystals. The underlying theory is directly applicable to more than one molecule as the central unit. However, such a system could not conveniently be handled in the current programme. On the other hand, it would be much more efficient to treat different molecules in different QM calculations. This is generally performed in other QM/MM approaches. It is not directly transferable to the ac-QM/MM model, as it assumes that a single unit can represent the whole crystal. However, it might be possible to generalise the approach. Then one could make the assumption that for different (i.e. not symmetry equivalent) molecules the ac-QM/MM model is individually applicable. Hence, one could calculate the contributions of all different molecule individually and sum these up. It would, therefore, be possible to save the efficiency gain from the use of symmetry relations, but beyond the initial idea of all molecules in the crystal to be the same.

To bridge the gap between experiment and theory, one desirable implementation to have is a vibrational analysis. The ac-QM/MM model might formally even be used to calculate lattice vibrations, since the required Hessian matrix elements are accessible (at least numerically). However, a simple approach for vibrations in molecular crystals⁹⁶ would fit very well to the ac-QM/MM model. It is assumed that, due to the stiffness of chemical bonds, one can apply the Einstein model of individually vibrating particles (molecules, in this case). Thus, one can restrict the calculation to the uniform vibration of all molecules. This is in line with the basic idea of the ac-QM/MM model. It would allow for the comparison with

experimental vibrational spectra, e.g. in order to distinguish packing motifs. Moreover, this would allow the estimation of enthalpies and free energies.

Bibliography

- ¹ C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101.
- ² H. Schenk, R. Peschar, *Radiat. Phys. and Chem.* **2004**, *71*, 829.
- ³ J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, J. Morris, *Pharm. Res.* **2001**, *18*, 859.
- ⁴ M. Bühl, T. van Mourik, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 634.
- ⁵ V. V. Nosenko, G. Y. Rudko, A. M. Yaremko, V. O. Yukhymchuk, O. M. Hreshchuk, *J. Raman Spectrosc.* **2018**, *49*, 559.
- ⁶ I. Georgieva, N. Trendafilova, N. Dodoff, D. Kovacheva, *Spectrochim. Acta, Part A* **2017**, *176*, 58.
- ⁷ Z. Fei, N. Kocher, C. J. Mohrschladt, H. Ihmels, D. Stalke, *Angew. Chem., Int. Ed.* **2003**, *42*, 783.
- ⁸ S. Haas, A. F. Stassen, G. Schuck, K. P. Pernstich, D. J. Gundlach, B. Batlogg, U. Berens, H.-J. Kirner, *Phys. Rev. B* **2007**, *76*, 115203.
- ⁹ C. M. Reddy, G. Rama Krishna, S. Ghosh, *CrystEngComm* **2010**, *12*, 2296.
- ¹⁰ J. Hey, D. M. Andrada, R. Michel, R. A. Mata, D. Stalke, *Angew. Chem., Int. Ed.* **2013**, *52*, 10365.
- ¹¹ S. Miertuš, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117.
- ¹² J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. Van Eijck, P. Verwer, D. E. Williams, *Acta Crystallogr., Sect. B* **2000**, *56*, 697.
- ¹³ W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. Van Eijck, P. Verwer, D. E. Williams, *Acta Crystallogr., Sect. B* **2002**, *58*, 647.
- ¹⁴ G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. Van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, P. Verwer, *Acta Crystallogr., Sect. B* **2005**, *61*, 511.
- ¹⁵ G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. Van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. Van De Streek, A. K. Wolf, B. Schweizer, *Acta Crystallogr., Sect. B* **2009**, *65*, 107.
- ¹⁶ D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta,

- S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti, I. K. Zhitkov, *Acta Crystallogr., Sect. B* **2011**, *67*, 535.
- ¹⁷ A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. D. Jr, A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, C. R. Groom, *Acta Crystallogr., Sect. B* **2016**, *72*, 439.
- ¹⁸ J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, G. K.-L. Chan, *Science* **2014**, *345*, 640.
- ¹⁹ F. Bloch, *Z. Phys.* **1929**, *52*, 555.
- ²⁰ P. J. Bygrave, N. L. Allan, F. R. Manby, *J. Chem. Phys.* **2012**, *137*, 164102.
- ²¹ P. Mörschel, M. U. Schmidt, *Acta Crystallogr., Sect. A* **2015**, *71*, 26.
- ²² R. Bjornsson, M. Bühl, *J. Chem. Theory Comput.* **2012**, *8*, 498.
- ²³ G. J. O. Beran, K. Nanda, *J. Phys. Chem. Letters* **2010**, *1*, 3480.
- ²⁴ H. Stoll, *Phys. Rev. B* **1992**, *46*, 6700.
- ²⁵ *ChemShell, a Computational Chemistry Shell. see www.chemshell.org.*
- ²⁶ P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1.
- ²⁷ E. Schrödinger, *Ann. Phys.* **1926**, *79*, 361.
- ²⁸ M. Born, R. Oppenheimer, *Ann. Phys.* **1927**, *389*, 457.
- ²⁹ V. Fock, *Z. Phys.* **1930**, *61*, 126.
- ³⁰ W. Pauli, *Z. Phys.* **1925**, *31*, 765.
- ³¹ J. C. Slater, *Phys. Rev.* **1929**, *34*, 1293.
- ³² C. Möller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.
- ³³ S. Grimme, *J. Chem. Phys.* **2003**, *118*, 9095.
- ³⁴ F. Coester, H. Kümmel, *Nucl. Phys.* **1960**, *17*, 477.
- ³⁵ J. Čížek, *J. Chem. Phys.* **1966**, *45*, 4256.
- ³⁶ K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, *Chem. Phys. Lett.* **1989**, *157*, 479.

- ³⁷ J. Pipek, P. G. Mezey, *J. Chem. Phys.* **1989**, *90*, 4916.
- ³⁸ A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.* **1985**, *83*, 735.
- ³⁹ R. A. Mata, H.-J. Werner, *Mol. Phys.* **2007**, *105*, 2753.
- ⁴⁰ S. Saebo/, W. Tong, P. Pulay, *J. Chem. Phys.* **1993**, *98*, 2170.
- ⁴¹ P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.
- ⁴² W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.
- ⁴³ S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- ⁴⁴ S. Rybak, B. Jeziorski, K. Szalewicz, *J. Chem. Phys.* **1991**, *95*, 6576.
- ⁴⁵ H. L. Williams, C. F. Chabalowski, *J. Phys. Chem. A* **2001**, *105*, 646.
- ⁴⁶ A. Heßelmann, G. Jansen, *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010.
- ⁴⁷ J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, J. G. Ángyán, *J. Chem. Phys.* **2006**, *124*, 154709.
- ⁴⁸ S. Suhai, *Phys. Rev. B* **1983**, *27*, 3506.
- ⁴⁹ P. Y. Ayala, K. N. Kudin, G. E. Scuseria, *J. Chem. Phys.* **2001**, *115*, 9698.
- ⁵⁰ C. Pisani, M. Busso, G. Capecchi, S. Casassa, R. Dovesi, L. Maschio, C. Zicovich-Wilson, M. Schütz, *J. Chem. Phys.* **2005**, *122*, 094113.
- ⁵¹ C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, D. Usvyat, *J. Comput. Chem.* **2008**, *29*, 2113.
- ⁵² F. Hummel, T. Tsatsoulis, A. Grüneis, *J. Chem. Phys.* **2017**, *146*, 124105.
- ⁵³ A. Hermann, P. Schwerdtfeger, *Phys. Rev. Lett.* **2008**, *101*, 183005.
- ⁵⁴ A. Hermann, P. Schwerdtfeger, *J. Chem. Phys.* **2009**, *131*, 244508.
- ⁵⁵ O. Bludský, M. Rubeš, P. Soldán, *Phys. Rev. B* **2008**, *77*, 092103.
- ⁵⁶ S. Tsuzuki, H. Orita, K. Honda, M. Mikami, *J. Phys. Chem. B* **2010**, *114*, 6799.
- ⁵⁷ A. D. Boese, J. Sauer, *Cryst. Growth Des.* **2017**, *17*, 1636.
- ⁵⁸ G. A. Dolgonos, O. A. Loboda, A. D. Boese, *J. Phys. Chem. A* **2018**, *122*, 708.
- ⁵⁹ E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, F. Weinhold, *GenNBO 5.9*, **2009**, Theoretical Chemistry Institute, University of Wisconsin, Madison.
- ⁶⁰ J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
- ⁶¹ V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, *J. Chem. Phys.* **2003**, *119*, 12129.
- ⁶² A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- ⁶³ A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, *123*, 154101.
- ⁶⁴ E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2005**, *123*, 024101.
- ⁶⁵ E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2006**, *124*, 174104.
- ⁶⁶ F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

- ⁶⁷ E. Baerends, D. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41.
- ⁶⁸ B. I. Dunlap, J. W. D. Connolly, J. R. Sabin, *J. Chem. Phys.* **1979**, *71*, 3396.
- ⁶⁹ J. L. Whitten, *J. Chem. Phys.* **1973**, *58*, 4496.
- ⁷⁰ R. Polly, H.-J. Werner, F. R. Manby, P. J. Knowles, *Mol. Phys.* **2004**, *102*, 2311.
- ⁷¹ K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283.
- ⁷² F. Neese, F. Wennmohs, A. Hansen, U. Becker, *Chem. Phys.* **2009**, *356*, 98.
- ⁷³ C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.* **1990**, *11*, 361.
- ⁷⁴ M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 09, Revision A.02*, Gaussian, Inc., Wallingford CT, 2016.
- ⁷⁵ H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, *MOLPRO, version 2012.1, a package of ab initio programs*, **2012**, see <http://www.molpro.net>.
- ⁷⁶ M. Schütz, G. Hetzer, H.-J. Werner, *J. Chem. Phys.* **1999**, *111*, 5691.
- ⁷⁷ G. Hetzer, M. Schütz, H. Stoll, H.-J. Werner, *The Journal of Chemical Physics* **2000**, *113*, 9443.
- ⁷⁸ C. Hampel, H. Werner, *J. Chem. Phys.* **1996**, *104*, 6286.
- ⁷⁹ C. Hampel, K. A. Peterson, H.-J. Werner, *Chem. Phys. Lett.* **1992**, *190*, 1.
- ⁸⁰ M. Schütz, H.-J. Werner, *The Journal of Chemical Physics* **2001**, *114*, 661.
- ⁸¹ J. D. Watts, J. Gauss, R. J. Bartlett, *J. Chem. Phys.* **1993**, *98*, 8718.
- ⁸² M. Schütz, H.-J. Werner, *Chem. Phys. Lett.* **2000**, *318*, 370.
- ⁸³ M. Schütz, *The Journal of Chemical Physics* **2000**, *113*, 9986.
- ⁸⁴ H.-J. Werner, F. R. Manby, P. J. Knowles, *The Journal of Chemical Physics* **2003**, *118*, 8149.
- ⁸⁵ M. Schutz, F. R. Manby, *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349.
- ⁸⁶ T. H. Dunning Jr., *J. Chem. Phys.* **1989**, *90*, 1007.
- ⁸⁷ F. Weigend, *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.
- ⁸⁸ F. Weigend, A. Köhn, C. Hättig, *J. Chem. Phys.* **2002**, *116*, 3175.

- ⁸⁹ A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- ⁹⁰ J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2011**, *25*, 1157.
- ⁹¹ K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell, *J. Comput. Chem.* **2010**, *31*, 671.
- ⁹² N. L. Allinger, Y. H. Yuh, J. H. Lii, *J. Am. Chem. Soc.* **1989**, *111*, 8551.
- ⁹³ J. W. Ponder, *Tinker - Software Tools for Molecular Design*, see <https://dasher.wustl.edu/tinker/>.
- ⁹⁴ J. Kästner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, P. Sherwood, *J. Phys. Chem. A* **2009**, *113*, 11856.
- ⁹⁵ D. C. Liu, J. Nocedal, *Math. Program.* **1989**, *45*, 503.
- ⁹⁶ A. Otero-de-la Roza, E. R. Johnson, *J. Chem. Phys.* **2012**, *137*, 054103.
- ⁹⁷ A. M. Reilly, A. Tkatchenko, *J. Chem. Phys.* **2013**, *139*, 024705.
- ⁹⁸ J. Moellmann, S. Grimme, *J. Phys. Chem. C* **2014**, *118*, 7615.
- ⁹⁹ S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, M. C. Payne, *Z. Kristallogr. - Cryst. Mater.* **2005**, *220*, 567.
- ¹⁰⁰ C. G. Broyden, *IMA J. Appl. Math.* **1970**, *6*, 76.
- ¹⁰¹ R. Fletcher, *Comput. J.* **1970**, *13*, 317.
- ¹⁰² D. Goldfarb, *Math. Comput.* **1970**, *24*, 23.
- ¹⁰³ D. F. Shanno, *Math. Comput.* **1970**, *24*, 647.
- ¹⁰⁴ S. Alexandratos, A. Streitwieser, H. F. Schaefer, *J. Am. Chem. Soc.* **1976**, *98*, 7959.
- ¹⁰⁵ K. C. Waterman, A. Streitwieser, *J. Am. Chem. Soc.* **1984**, *106*, 3138.
- ¹⁰⁶ C. J. Schürmann, T. L. Teuteberg, C. Stückl, R. Herbst-Irmer, R. A. Mata, D. Stalke, *On the Brink to Radiation Damage: Observation of the homolytic Se-C Bond Cleavage in x- and UV-irradiated (BzSe)₂ Crystals*, in preparation.
- ¹⁰⁷ P. I. Sampath, *J. Chem. Phys.* **1966**, *45*, 3519.
- ¹⁰⁸ J. J. Windle, A. K. Wiersema, A. L. Tappel, *J. Chem. Phys.* **1964**, *41*, 1996.
- ¹⁰⁹ R. A. Jackson, R. Moosavi, *J. Chem. Soc., Perkin Trans. 2* **1992**, 885.
- ¹¹⁰ S. Zheng, Q. Tang, J. He, S. Du, S. Xu, C. Wang, Y. Xu, F. Lin, *J. Chem. Inf. Model.* **2016**, *56*, 811.
- ¹¹¹ S. Grimme, *J. Chem. Theory Comput.* **2014**, *10*, 4497.
- ¹¹² P. Maurer, A. Laio, H. W. Hugosson, M. C. Colombo, U. Rothlisberger, *J. Chem. Theory Comput.* **2007**, *3*, 628.
- ¹¹³ L. Sun, L. Yang, Y.-D. Zhang, Q. Shi, R.-F. Lu, W.-Q. Deng, *J. Comput. Chem.* **2017**, *38*, 1991.

