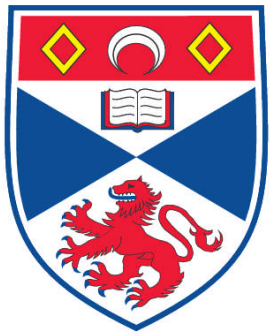
ENVIRONMENTAL EFFECTS IN QUANTUM CHEMISTRY: QM/MM STUDIES OF STRUCTURES, NMR PROPERTIES AND REACTIVITIES IN EXTENDED SYSTEMS

Ragnar Björnsson

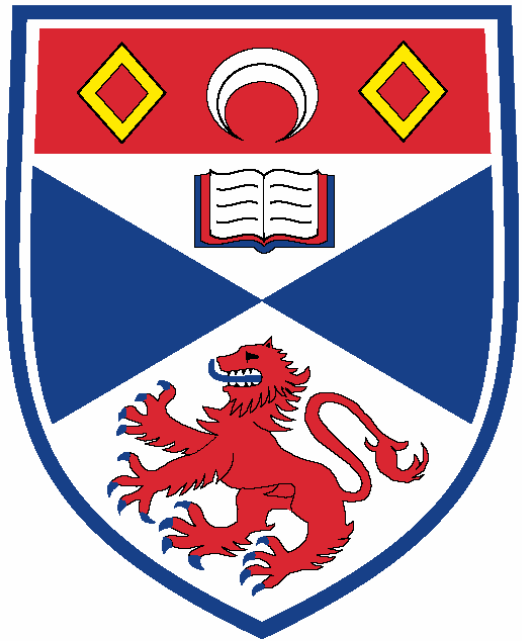
A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews



2012

Environmental effects in quantum chemistry: QM/MM studies of structures, NMR properties and reactivities in extended systems

Ragnar Björnsson



A thesis submitted to the University of St. Andrews for the degree of DOCTOR OF PHILOSOPHY

School of Chemistry

March 2012

Declarations

I, Ragnar Björnsson, hereby certify that this thesis, which is approximately 60000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in October 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2012.

Date Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date Signature of supervisor

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date Signature of candidate

Signature of supervisor

"It is also a good rule not to put too much confidence in experimental results until they have been confirmed by theory."
Sir Arthur Eddington

Acknowledgements

I'd like to thank:

- My supervisor Michael Bühl for excellent support and many good discussions during my PhD, in particular for allowing my PhD project to spin out in many different directions that has given me experience in many different branches of computational chemistry.

- EaStCHEM for funding my PhD project and the EaStCHEM Research Computing Facility for the available computer cluster facilities.

- Herbert Früchtl for various computational support and discussions.

- Tanja van Mourik for discussions about the 3F-GABA project and computational chemistry in general.

- Tom Keal and Paul Sherwood in Daresbury for much remote support with the Chemshell software.

- Uli Schwarz-Linek for many good discussions about the isopeptide bond project.

- Alex and the staff of Russel Hotel for good food & service in their excellent quiet pub, stimulating many ideas of this thesis.

- Lab coworkers in room 208 and 150 of the Purdie building:
  Lilianna, Lukasz, Jie, Nicolas, James, Alex, Luke, Neetika, Rosie, Leo, John and Lazaros for friendship and great atmosphere in the office and the many excellent trips to the pub.

- My parents for their many trips to Scotland visiting me and other family members for support over the years.

Publications resulting from this thesis

Modelling Molecular Crystals by QM/MM: Self-Consistent Electrostatic Embedding for Geometry Optimisations and Molecular Property Calculations in the Solid
Bjornsson, R.; Bühl, M. *J. Chem. Theory Comput.*, **2012**, *8*, 498-508.

Modelling zwitterions in solution: 3-fluoro-γ-aminobutyric acid (3F-GABA)
Cao, J.; Bjornsson, R.; Bühl, M.; Thiel, W.; van Mourik, T. *Chem. Eur. J.*, **2012**, *18*, 184-195.

^{51}V NMR parameters of $VOCl_3$: Static and dynamic density functional study from the gas-phase to the bulk
Bjornsson, R.; Früchtl, H.; Bühl, M. *Phys. Chem. Chem. Phys.*, **2011**, *13*, 619-627.

NMR and Theoretical Analysis of a Spontaneously Formed Lys-Asp Isopeptide Bond
Hagan, R. M.; Bjornsson, R.; McMahon, S. A.; Schomburg, B.; Braithwaite, V.; Bühl, M.; Naismith, J. H.; Schwarz-Linek, U. *Angew. Chem. Int. Ed.*, **2010**, *45*, 8421-8425.

Electric Field Gradients of Transition Metal Complexes from Density Functional Theory: Assessment of Functionals, Geometries and Basis sets.
Bjornsson, R.; Bühl, M. *Dalton. T.*, **2010**, *39*, 5319-5324.

Abstract

Computational modelling of chemical systems is most easily carried out in the vacuum for single molecules. Accounting for environmental effects accurately in quantum chemical calculations, however, is often necessary for computational predictions of chemical systems to have any relevance to experiment.

This PhD thesis focuses on accounting for environmental effects in quantum chemical calculations by quantum mechanics/ molecular mechanics (QM/MM) approaches, taking on diverse examples from the solid state, the liquid phase and the protein environment. The methods are applied to compute a variety of properties from transition metal NMR properties of molecular crystals and enzymes, via conformational properties of zwitterions in aqueous solution, to an intramolecular amidation reaction in peptides.

Chapter 3 concerns QM/MM calculations of molecular properties in the solid state, both molecular crystals and metalloenzymes, with a focus on transition metal chemical shift and EFG properties. We demonstrate that solid-state effects on such properties in molecular crystals can be accounted for by a simple general black-box QM/MM approach. We also describe preliminary QM/MM calculations of ^{51}V anisotropic NMR properties for a vanadium-dependent enzyme.

In Chapter 4 the focus is on solvent effects on the conformational preference of a small zwitterionic molecule, 3F- γ-aminobutyric acid (3F-GABA), calculated using QM/MM molecular dynamics simulations. NMR spin-spin coupling constants in solution are also calculated. Our simulations highlight the difficulty of accounting for solvation effects well enough to achieve agreement with experimental observations.

Chapter 5 concerns the reaction mechanism of an intramolecular amidation reaction in a bacterial peptide, predicted by QM/MM calculations. We predict a reaction mechanism that accounts well for the experimental observations both for the wild-type and mutants.

We demonstrate that environmental effects can often be satisfactorily accounted for by QM/MM approaches, thus helping to bridge the gap between theory and experiment.

Table of contents

Abbreviations

DFT	Density Functional Theory
EFG	Electric Field Gradient
ESP	Electrostatic Potential
GABA	γ-aminobutyric acid
HF	Hartree-Fock
MAE	Mean absolute error
MD	Molecular dynamics
ME	Mean error
MaxE	Maximum error
MW	Microwave
MM	Molecular Mechanics
NPA	Natural Population Analysis
NMR	Nuclear Magnetic Resonance
NQCC	Nuclar Quadrupole Coupling Constant
PCM	Polarisable Continuum Model
PES	Potential Energy Surface
QM	Quantum Mechanics
QM/MM	Quantum Mechanics / Molecular Mechanics
SCF	Self Consistent Field
VCPO	Vanadium chloroperoxidase
WFT	Wavefunction Theory

Chapter 1. Introduction

Computational chemistry is gradually becoming an important subdiscipline of chemistry. By using approximations to the fundamental equations of quantum mechanics, properties of molecules can be calculated with computers, even starting from nothing but the chemical formula. The quantum chemical methods that are in use, however, suffer from a severe scaling problem in terms of computational cost as the molecules get larger. Methods based on wavefunction theory (WFT) and density functional theory (DFT) can nowadays easily be used for calculations on molecules up to tens and up to a few hundred atoms, respectively.[1,2] Much interesting chemistry that is the focus of today, however, takes place on a larger scale, e.g. in the field of biochemistry and materials chemistry. Classical molecular mechanics can deal with such an increase in system size but only after massive parameterisation against experiment and at the expense of predictive power.

Additionally, the chemistry that takes place in the condensed phases introduces environmental effects that can require a substantial increase in the number of atoms. Often these effects are ignored, and gas-phase models of the real chemical systems are calculated instead, thus potentially introducing a severe model error to the calculation. A major challenge for molecular modelling is to move beyond gas-phase models and and to properly take into account the relevant environmental effects. Reliable and affordable methods are needed for this to be possible.

Modelling environmental effects and large chemical systems with quantum chemical methods is the focus of this work. Three types of environmental effects will be discussed and our attempts to model them will be described, with a chapter devoted to each of them.

Solid-state effects on molecular properties in molecular crystals

The solid phase can be rather diverse, ranging from ionic crystals, metal oxides, zeolites, polymers and molecular crystals. In molecular crystals, small organic or inorganic compounds have crystallised due to attractive intermolecular interactions and it is often the solid state where the structure determination of the compound takes place, for example by X-ray crystallography, solid-state NMR, Mössbauer, EPR, IR and Raman spectroscopies. While it is usually the single molecule properties that are of interest to the chemist, the solid state introduces crystal-packing effects due to the many intermolecular interactions between molecules in a crystal and a polar crystal environment will inevitably influence some molecular properties to an unknown degree. Additionally, counter-ions and even crystallised solvent may significantly alter the picture of the environment surrounding each molecule, compared to a single molecule in the gas-phase. This creates a problem when modelling molecules for comparison with experiment or interpretation of solid-state experimental data as any solid-state effect is effectively lost when only a single molecule is subjected to quantum chemical calculations.

Solvent effects

The liquid phase creates similar problems. Many reactions in organic chemistry have well established solvent effects that affect the reaction kinetics. There are also molecules that adopt a different structure compared to the gas-phase or the solid

phase. The disorder and flexibility of the liquid phase also considerably complicates the modelling.
Much effort has been spent over the years on approximating the liquid phase by implicit (continuum) solvation models.[3] While these methods have been remarkably successful, they fail for many systems and it is clear that for improved accuracy, the explicit nature of the solvent must eventually be taken into account. Explicit modelling of the solvent in quantum chemical calculations, however, not only increases the computational cost substantially; due to the increase in degrees of freedom of the system, multiple local minima become available, creating problems for traditional conformational analysis on a potential energy surface.

The effect of the protein environment

Modelling reactions taking place in proteins shares many of the aforementioned problems of the solid and solvent phases. While the protein reaction can often be qualitatively well described by only a few atoms, the protein introduces an environmental effect that directly affects the reaction kinetics, leading to the dramatic efficiency of enzyme catalysis compared to the reaction taking place in the gas or solvent phase. The flexibility of the protein or the constrained motion of some residues in the active site can be directly responsible for how a reaction proceeds in a protein environment. This information is often lost when, e.g. only the active site of an enzyme is modelled in the vacuum.

Quantum mechanics/molecular mechanics (QM/MM) [4-6] is an approach to take large molecular systems into the realm of quantum chemistry and/or to explicitly account for environmental effects on the properties of a molecule or a chemical reaction. The molecular system is divided into a QM region (where the relevant chemistry takes place) and an MM region (the surroundings), and a quantum chemical method is applied to the former region and classical force fields to the latter. From the point of a typical quantum chemical calculation, this introduces roughly two main additional approximations: (i) the way the MM region is described and (ii) the way the molecular system is divided into the QM region and the MM region and how the regions interact with each other.

This PhD thesis focuses on accounting for environmental effects in quantum chemical calculations by QM/MM approaches, taking on diverse examples from solid, liquid and protein phases, such as transition metal NMR properties of molecular crystals, conformational properties of zwitterions in aqueous solution and intramolecular amidation reactions in peptides. Chapter 3 concerns QM/MM calculations of molecular properties in the solid state, both molecular crystals and metalloenzymes with a focus on transition metal chemical shift and EFG properties. In Chapter 4 the focus is on solvent effects on a small zwitterionic molecule, 3F-GABA, calculated using QM/MM molecular dynamics simulations. NMR coupling constants in solution are also calculated. Chapter 5 concerns the reaction mechanism of an intramolecular amidation reaction in a bacterial peptide, calculated by QM/MM calculations.

Chapter 2. Theoretical background

2.1 Many-electron quantum mechanics

The physical laws responsible for chemical phenomena are known and are rooted in quantum mechanics. The non-relativistic Schrödinger equation describes how the wavefunction, Ψ, of a quantum particle changes with respect to time (t):

$$i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} = H\Psi(\mathbf{r},t) \tag{1}$$

where the Hamilton operator is:

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial\mathbf{r}^2} + V(\mathbf{r},t) \tag{2}$$

and *m* is the mass of the particle, *ħ* is the reduced Planck's constant and *V*(**r**,*t*) is potential dependent on both the spatial coordinates **r** of the particle and time *t*. The wavefunction describes the motion and location of the particle and contains everything that can be known about the system (one of the postulates of quantum mechanics).

Operators are used (such as the Hamiltonian) to get information from the wavefunction. When a time-independent potential, *V*(**r**), is defined, the time-independent form of the Schrödinger equation is used to describe the stationary states of the wavefunction, *ψ*. It is an eigenvalue equation relating the energy *E* (the eigenvalue), the wavefunction *ψ* (the eigenfunction) and the Hamilton operator *H*:

$$H\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{3}$$

This equation can in principle be solved to yield the ground and excited state wavefunctions for any quantum system with a time-independent potential. The wavefunction itself does not, however, have a direct interpretation or connection to experiment. Instead, the square of the wavefunction is interpreted as a probability density, yielding a probability that the particle will be there when multiplied with some volume element d**r**.

$$\rho = |\psi(\mathbf{r})|^2 d\mathbf{r} \tag{4}$$

In a many-particle system like an atom or molecule, however, the Schrödinger equation becomes very complex due to the interdependency of nuclear and electronic motion (no particle is moving independently) and becomes analytically insolvable as can be seen from the form the Hamiltonian takes for a system containing nuclei and electrons :

$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i\sum_k \frac{e^2 Z_k}{4\pi\epsilon_0 r_{ik}} + \sum_{i<j}\frac{e^2}{4\pi\epsilon_0 r_{ij}} + \sum_{k<l}\frac{e^2 Z_k Z_l}{4\pi\epsilon_0 r_{kl}} \tag{5}$$

where *i* and *j* are the indices of the electrons and *k* and *l* are the indices of the nuclei and the Laplace operator is:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{6}$$

The *m* stands for mass, *r* for inter-particle distance, *e* the elementary charge, *Z* the nucleus charge and $4\pi\varepsilon_0$ is the vacuum permittivity.
The first and second terms of Eq. 5 describe the kinetic energy of the electrons and nuclei, the third term describes the potential energy between the electrons and nuclei, and the fourth and fifth terms describe the potential energy of electron-electron and nuclei-nuclei interactions. A crucial approximation, that greatly simplifies practical calculations, is the separation of nuclear and electronic motion. This is the Born-Oppenheimer approximation which can be justified by the fact that nuclei and electrons move on different timescales due to their mass differences. This enables one to solve the electronic Schrödinger equation separately, which ignores the kinetic energy of the nuclei and only includes the nuclear-nuclear potential energy as a constant classical energy term (for each set of nuclear positions):[1]

$$(H_{elec} + V_{NN})\psi_{elec} = E_{elec}\psi_{elec} \tag{7}$$

$$H_{elec} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_i \sum_k \frac{e^2 Z_k}{4\pi\epsilon_0 r_{ik}} + \sum_{i<j} \frac{e^2}{4\pi\epsilon_0 r_{ij}} \tag{8}$$

$$V_{NN} = \sum_{k<l} \frac{e^2 Z_k Z_l}{4\pi\epsilon_0 r_{kl}} \tag{9}$$

The nuclear coordinates are here parameters, not variables, and the electronic Schrödinger equation is solved for one set of nuclear coordinates that enter into the second term of Eq. 8 and into Eq. 9. Solving the Schrödinger equation for different sets of nuclear coordinates gives a Born-Oppenheimer potential energy surface (PES) which is of great conceptual value in chemistry. The Born-Oppenheimer approximation effectively strips the nuclei of their quantum nature, resulting in simple classical point charges (and additionally without mass or kinetic energy in Eqs. 7-9).

The nuclear kinetic energy which is ignored in the electronic Schrödinger equation can later be recovered by solving a nuclear Schrödinger equation:

$$\left(-\sum_k \frac{1}{2m_k}\nabla_k^2 + V(\mathbf{r_N}) \right)\psi_{nuc} = E\psi_{nuc} \tag{10}$$

where the electronic interaction enters as a potential energy term $V(\mathbf{r_N})=E_{\text{elec}}+V_{\text{NN}}$ dependent on nuclear coordinates. The electronic potential has to be approximated, usually as a harmonic potential based on the electronic solution. This enables one to recover the zero-point energy as well as temperature-dependent vibrational motion that can be used to calculate energies, entropies and free energies at any temperature using the equations of statistical mechanics. Another way to recover the nuclear kinetic energy is by molecular dynamics simulations as will be discussed later.

2.2 Solving the Schrödinger equation

2.2.1 The variational principle

While the Born-Oppenheimer approximation greatly simplifies the interdepency-of-particles problem, the electron-electron interaction term is still left, meaning that the each electron is dependent on all other electrons. Any many-body problem containing 3 or more interdependent particles is analytically insolvable and thus in order to do practical calculations, more approximations must be invoked.

Many of the approximations that have been invoked over the years, make use of the variational principle. It states that if a system with Hamiltonian H is described by an allowed trial wavefunction ψ_{trial} then the following equation holds:

$$\frac{\int \psi_{trial}^{*} H \psi_{trial} d\mathbf{r}}{\int \psi_{trial}^{*} \psi_{trial} d\mathbf{r}} = E_{trial} \geq E_0 \tag{11}$$

where ψ^* is the complex conjugate of ψ, E_{trial} is the energy of the trial wavefunction and E_0 is the energy of the exact groundstate wavefunction. The inequality thus guarantees that the energy of the trial wavefunction will always be higher than the real energy of the real wavefunction except for the case that the trial wavefunction is equal to the real wavefunction.
The variational equation thus suggests a route towards approaching the real wavefunction: either by trying all possible wavefunctions allowed by the number of particles or to somehow systematically improve upon the trial wavefunction until its energy cannot get lower.

2.2.2 Hartree-Fock theory

Even though the Born-Oppenheimer approximation and the variational principle together provide a route towards finding the electronic wavefunction we cannot search through all possible trial wavefunctions having the correct number of electrons as there are simply too many possibilities and in fact we need a form for the wavefunction on which the Hamiltonian can operate.
However, approximating the *N*-electron wavefunction as a linear combination of 1-electron wavefunctions provides a convenient starting point. Such an allowed wavefunction is the Slater determinant which is an antisymmetric product of 1-electron wavefunctions that obeys the postulate that electrons are indistinguishable and cannot be in the same state (Pauli exclusion principle) :

$$\psi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \dots & \phi_N(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \dots & \phi_N(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\mathbf{x}_N) & \phi_2(\mathbf{x}_N) & \dots & \phi_N(\mathbf{x}_N) \end{vmatrix} \tag{12}$$

The $\phi_i(\mathbf{x})$ are the 1-electron wavefunctions, called spin orbitals i.e. products of both a spatial orbital and a spin function. The spin orbitals are functions of both spatial and spin degrees of freedom, together denoted $(\mathbf{x})$.

Having now a way to build up a wavefunction for any number of electrons we can use the Born-Oppenheimer Hamiltonian and the variational principle to minimise the wavefunction to yield the ground state wavefunction and the ground state energy. Atomic units will be adopted from now on, meaning that the electron mass, the elementary charge, Planck's constant ħ and the vacuum permittivity $4\pi\varepsilon_0$ will be set to unity.

This leads to the Hartree-Fock energy equation:

$$E_{HF} = \langle \psi_{SD} | H_{elec} + V_{NN} | \psi_{SD} \rangle = \sum_i \langle \phi_i | h | \phi_i \rangle + \frac{1}{2} \sum_i \sum_j \langle \phi_i \phi_i | \phi_j \phi_j \rangle - \langle \phi_i \phi_j | \phi_j \phi_i \rangle + V_{NN} \tag{13}$$

where we have introduced the Dirac short hand notation. The individual terms of eq. 13 are:

$$\langle \phi_i | h | \phi_i \rangle = \int \phi_i^*(\mathbf{x}) \left(-\frac{1}{2}\nabla^2 - \sum_k \frac{Z_k}{r_{ik}} \right) \phi_i(\mathbf{x}) d\mathbf{x} \tag{14}$$

$$\langle \phi_i \phi_i | \phi_j \phi_j \rangle - \langle \phi_i \phi_j | \phi_j \phi_i \rangle = \int \int |\phi_i(\mathbf{x}_1)|^2 \frac{1}{r_{12}} |\phi_j(\mathbf{x}_2)|^2 d\mathbf{x}_1 d\mathbf{x}_2 - \int \int \phi_i(\mathbf{x}_1) \phi_j^*(\mathbf{x}_1) \frac{1}{r_{12}} \phi_j(\mathbf{x}_2) \phi_i^*(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{15}$$

Eq. 14 is an integral that describes the kinetic energy and the nucleus-electron interaction and is a 1-electron equation. Eq. 15 shows two double integrals that describe the electron-electron repulsion. The first term is the Coulomb integral (often denoted J_{ij}) that represents the electrostatic repulsion of two smeared out charge clouds while the second term is the exchange integral (often denoted K_{ij}) that is impossible to interpret classically, but arises due to the antisymmetry of the Slater determinant that was enforced.

As we have built up the many-electron wavefunction from 1-electron wavefunctions (orbitals), our variational freedom now comes from varying the orbitals (with the constraint that the orbitals must remain orthonormal) and thus we can minimise the whole energy expression by choosing the best orbitals. This is achieved by solving the Hartree-Fock equations:[7]

$$f \phi_i = \epsilon_i \phi_i \tag{16}$$

for each electron where ε_i are the eigenvalues to the Fock operator f which is defined as:

$$f = -\frac{1}{2}\nabla_i^2 - \sum_k \frac{Z_k}{r_{ik}} + V_{HF} \tag{17}$$

where V_{HF} is the Hartree-Fock potential which is an average repulsive potential (defined by Coulomb and exchange operators, not shown) due to all the electrons except for the one which is being solved. As the Fock operator both operates on the orbitals and depends on the orbitals through the Hartree-Fock potential, the Hartree-Fock equations have to be solved self-consistently by first guessing the initial orbitals and then iteratating until the orbitals stop changing. Once self-consistent orbitals are available, the Hartree-Fock energy equation (Eq. 13) gives the Hartree-Fock total energy while the orbital energies can be interpreted as approximate ionisation energies of electrons from the orbitals that they occupy.

2.2.3 Post Hartree-Fock approaches

Hartree-Fock theory is remarkably successful and recovers ~99.9 % of the exact ground state energy of a small- to medium-sized molecule.[8] Unfortunately, the neglected fraction is still very large considering the energy scale of chemical reactions for example. Due to error cancellations when calculating energy differences, chemically meaningful results can often still be obtained, however, it turns out that the neglected energy is crucial to describe most chemical phenomena accurately. Electron correlation is usually defined as this neglected energy by the Hartree-Fock approximation and is hence the difference between the true Born-Oppenheimer energy and the Hartree-Fock energy. The effect due to electron correlation can be roughly described as the instantaneous electron-electron interaction that is missing from the Hartree-Fock description.

To go beyond Hartree-Fock theory and account for electron correlation, one strategy has been to create a wavefunction that consists of a linear combination of Slater determinants (Configuration Interaction). Starting from the minimised Hartree-Fock determinant, excited determinants are derived by performing allowed excitations of the electrons in the Hartree-Fock determinant and the linear combination is minimised by the variational principle. Ultimately, this strategy can lead to the exact Born-Oppenheimer electronic energy of any molecule as long as all possible excitations are performed, i.e. Full Configuration Interaction. This is only possible for the smallest of molecules. A similar strategy, coupled cluster theory, has become popular to calculate correlated wavefunctions that improve upon Hartree-Fock wavefunctions. Usually only a few types of excitations can be afforded which leads to the CCSD and CCSD(T) approximations (coupled cluster singles doubles perturbative triples). The CCSD(T) method is often referred to as the gold standard of quantum chemistry as it is a highly reliable and accurate method, especially for closed-shell organic molecules. It is very expensive compared to the Hartree-Fock method, however.
Methods based on these approaches will be referred to as post-HF methods or methods based on wavefunction theory (WFT).

2.2.4 Density functional theory

The correlated wavefunction approach systematically improves the wavefunction which, however, leads to very complicated methods that become too expensive to use for most molecules. Density functional theory, on the other hand, takes a different approach to solving the Schrödinger equation as it establishes the electron density as the quantity to derive the ground state energy from instead of the many-electron wavefunction.

The Born-Oppenheimer Hamiltonian (Eqs. 8 and 9) for any molecule is fully determined by the number of electrons N and the electron-nucleus potential υ (which is specified by nuclear positions and nuclear charges) and leads to a unique ground state Born-Oppenheimer energy and associated wavefunction of a molecule. The correlated wavefunction approach of the previous chapter is an obvious way towards the ground state energy (and wavefunction) as we have seen, but what if one could avoid the calculation of the complicated wavefunction? The electron density, $\rho(\mathbf{r})$, can be defined as a multiple integral of the square of the wavefunction over the spin coordinates of all electrons and over all but one spatial coordinate.[7]

$$\rho(\mathbf{r}) = N \int ... \int |\psi(\mathbf{r})|^2 ds d\mathbf{x_2} ... d\mathbf{x_N} \tag{18}$$

It gives the probability that any of the *N* electrons of the system of any spin is found within the volume element and is a simple function of the spatial coordinates. The electron density is experimentally observable, from X-ray diffraction for instance. This simple quantity turns out to contain all the necessary information: the number of electrons, *N*, can be determined by integrating over $\rho(\mathbf{r})$, and the nuclear positions and charges show up as cusps of $\rho(\mathbf{r})$.[8] That the nuclear positions and charges i.e. the electron-nucleus potential υ can be uniquely deduced from $\rho(\mathbf{r})$ is the subject of the first Hohenberg-Kohn theorem.[9]
It should thus be possible to come up with an energy functional (a functional is a special type of function whose argument is another function) whose argument is the electron density and determines the ground state energy. The energy functional would have to account for the electronic kinetic energy, the nucleus-electron interaction energy and the electron-electron interaction energy (the trivial nucleus-nucleus repulsion term is ignored in all following equations) in some way:

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + E_{Ne}[\rho(\mathbf{r})] + E_{ee}[\rho(\mathbf{r})] \tag{19}$$

where $T[\rho(\mathbf{r})]$ is the kinetic energy term, $E_{Ne}[\rho(\mathbf{r})]$ is the electron-nucleus term and $E_{ee}[\rho(\mathbf{r})]$ is the electron-electron interaction term, all expressed in terms of $\rho(\mathbf{r})$. The $E_{Ne}[\rho(\mathbf{r})]$ term is easy to express as a functional of the density, but it is far from obvious how to express $T[\rho(\mathbf{r})]$ and $E_{ee}[\rho(\mathbf{r})]$ in terms of $\rho(\mathbf{r})$.

Another problem concerns finding the ground state density with such an expression instead of just any density. Luckily, the second Hohenberg-Kohn theorem[9] provides density functional theory with its own variational principle which states that any trial density $\rho_{trial}(\mathbf{r})$ will always be higher in energy (or equal) than the real $\rho_{exact}(\mathbf{r})$ and thus Eq. 19 (if it can be expressed) can simply be tried for a number of trial densities or somehow minimised by systematically improving the trial density.

A simple expression of the $T[\rho(\mathbf{r})]$ kinetic energy term has been the subject of various attempts, some dating back to the early days of quantum mechanics, e.g. the Thomas-Fermi model.[10,11] These attempts, however, have not resulted in useful kinetic energy functionals though there has been recent progress on this front.[12]

In order to circumvent this problem some arguments from Hartree-Fock theory can be borrowed. By considering a fictitious system of non-interacting particles described by a Slater determinant of 1 electron orbitals, a non-interacting trial density $\rho_{KS}(\mathbf{r})$ can be constructed, as suggested by Kohn and Sham:[13]

$$\rho_{KS}(\mathbf{r}) = \sum_i \int |\phi_i(\mathbf{x})|^2 ds \tag{20}$$

By requiring $\rho_{KS}(\mathbf{r})$ to be equal to $\rho(\mathbf{r})$, one can still work with the exact equations of DFT and Eq. 19 can be rewritten as:

$$E[\rho(\mathbf{r})] = T_{KS}[\rho(\mathbf{r})] + E_{Ne}[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \tag{21}$$

where the kinetic energy term is defined as:

$$T_{KS}[\rho(\mathbf{r})] = -\frac{1}{2}\sum_i \langle\phi_i|\nabla^2|\phi_i\rangle \tag{22}$$

the electron-nucleus term is:

$$E_{Ne}[\rho(\mathbf{r})] = -\sum_k Z_k \int \rho(\mathbf{r}) r_k^{-1} d\mathbf{r} \tag{23}$$

and the Coulomb term is:

$$J[\rho(\mathbf{r})] = \frac{1}{2}\int\int \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) r_{12}^{-1} d\mathbf{r}_1 d\mathbf{r}_2 \tag{24}$$

where we have split the previous $E_{ee}[\rho(\mathbf{r})]$ term into a trivial Coulomb interaction term $J[\rho(\mathbf{r})]$ (calculated as in Hartree-Fock theory) and a new $E_{XC}[\rho(\mathbf{r})]$ term. While it may not be obvious, no approximation has been made. The $E_{XC}[\rho(\mathbf{r})]$ term, however, has become a rather complicated term, which must now account for a correction to the kinetic energy (not accounted for by the $T_{KS}[\rho(\mathbf{r})]$ term) as well as electron-electron repulsion not accounted for by the smeared charge cloud term $J[\rho(\mathbf{r})]$. Assuming an expression for $E_{XC}[\rho(\mathbf{r})]$ can be given, we can next use the Hohenberg-Kohn variational principle to solve the Kohn-Sham single particle equations to minimise energy w.r.t. the density as it is now expressed in orbitals:

$$\left(-\frac{1}{2}\nabla^2 + \upsilon_{eff}(\mathbf{r})\right)\phi_i(\mathbf{x}) = \epsilon_i\phi_i(\mathbf{x}) \tag{25}$$

where the effective potential, $\upsilon_{eff}(\mathbf{r})$ contains the nucleus-electron, Coulomb and exchange-correlation potentials.

The Kohn-Sham DFT equations are very similar to the Hartree-Fock equations and are also solved self-consistently and by the same type of computational techniques. The main difference is the absence of a *K* exchange term (which we will come back to) and the occurrence of an exchange-correlation term. The exchange-correlation energy can be thought of as the mutual avoidance of electrons due to the "personal space" (exchange-correlation hole) of each electron.

If an exact expression of the exchange-correlation term, $E_{XC}[\rho(\mathbf{r})]$, would be given, then solving the Kohn-Sham equations would lead to the exact solution to the Born-Oppenheimer Schrödinger equation. There is, however, no way of knowing this exact expression and educated guesses of its form must be used instead, which means that in practice DFT is always an approximate theory that suffers from the fact that its approximations are not easily systematically improvable unlike approximations based on WFT theory.

Theoretical developments of density functional theory for the last decades have thus been focused on "guessing" the exchange-correlation term, utilising various clues about its nature. This involves finding forms for the exchange and correlation energy that correctly describe the energy or density of model systems whose properties are well or exactly known. One of these model systems is the uniform electron gas, jellium whose correlation energy is well known from almost exact quantum Monte Carlo calculations (the exchange energy is exactly known). The local density

approximation (LDA) is a constituent of almost all exchange-correlation functionals in common use, and it is partly based on deriving a functional form (involving $\rho(\mathbf{r})$) that describes the properties of jellium well. Different exchange and correlation functionals based on the LDA approximation have been derived.[14-16] It turns out that the LDA approximation results in very inaccurate DFT energies when applied to molecules. In order to go beyond the LDA approximation, one strategy is to use the gradient of $\rho(\mathbf{r})$ as an ingredient in increasingly complex formulas for exchange and correlation respectively. The generalized gradient approximation (GGA) is the most successful gradient corrected extension of LDA which comes in numerous different varieties nowadays. A common gradient corrected exchange functional is the one by Becke (B88 or B)[17]. It is often combined with various different correlation functionals such as LYP[18] and P86[19]

Many different strategies have been employed in deriving other exchange-correlation functionals including better LDA approximations, density-gradient expansion, constraint satisfaction, modelling of the exchange-correlation hole, empirical fits and mixing in Hartree-Fock exchange.[20]

The mixing in of Hartree-Fock exchange (from Eq. 15) as first suggested by Becke[21] has been a particularly successful strategy, resulting in the very popular hybrid functionals that include both an expression of DFT exchange (usually one of the GGA exchange functionals) and Hartree-Fock exchange from Hartree-Fock theory, the two terms scaled empirically to avoid double counting. The inclusion of Hartree-Fock exchange introduces an explicit orbital-dependent term that is more expensive to compute than the other exchange-correlation terms, making hybrid functionals more expensive than LDA and GGA functionals. The very popular hybrid functional B3LYP is a 20 % Hartree-Fock exchange hybrid and is defined as follows:[22]

$$E^{B3LYP} = E_{XC}^{LDA} + 0.2(E_X^{HF} - E_X^{LDA}) + 0.72(E_X^{B88} - E_X^{LDA}) + 0.81(E_C^{LYP} - E_C^{LDA}) \quad (26)$$

Other hybrid functionals are defined in a similar way, with different exchange and correlation functionals used and a different amount of Hartree-Fock exchange. Hybrid functionals have been found to be the most successful functionals for general maingroup thermochemistry.

The reason for the success of the hybrid-DFT approach seems to partly stem from the self-interaction problem that plagues many LDA and GGA functionals and which particularly affects the accuracy of reaction barrier heights. The self-interaction problem comes from the Coulomb term $J[\rho(\mathbf{r})]$ of the DFT energy expression where an interaction of the electron with itself is calculated like in Hartree-Fock theory. However, in Hartree-Fock theory, the calculated Coulomb and exchange terms cancel exactly, removing this self-interaction, while in DFT there is no such full cancellation term. Mixing in a fraction of Hartree-Fock exchange seems to reduce the self-interaction problem of DFT with larger amounts reducing it more and removing it completely with 100 % Hartree-Fock exchange. However, this is no panacea as correlation functionals are generally incompatible with 100 % Hartree-Fock exchange[22] and most successful hybrids for reaction energies, for example, turn out to be 20-30 % hybrids, while the most accurate functionals for barrier heights are 40-60 % hybrids which reduces the accuracy for reaction energies etc. More recently, range-separated hybrids have been developed where the Hartree-Fock exchange term is no longer scaled by a constant but is evaluated differently based on the distance between electrons. This has resulted in increased flexibility of the hybrid functional form and special functionals with 100 % long-range Hartree-Fock exchange have been very

useful for predicting excited-state properties where the self-interaction error is prominent.[23]

Many recent functionals are of the meta-GGA type where another orbital-dependent term is added to the exchange-correlation functional in the form of the kinetic energy density. This gives additional flexibility to the functional form, allowing additional constraints (of the exact functional) to be satisfied. Hybrid meta-GGA functionals have also become common and a highly successful functional with a very complicated functional form and 35 empirical parameters that have been fitted to diverse data sets, is the M06-2X functional[24], containing 54 % Hartree-Fock exchange.

The ability and/or failure of common DFT approximations to describe dispersion has been discussed in the literature extensively in the last few years. It has been realized that most approximations show a completely repulsive potential energy curve for ideal gas dimers and that the failure to account for dispersive interactions has a considerable effect on general thermochemistry of main group compounds and grows in magnitude as larger molecules are calculated.[25] Many successful approaches to account for dispersive interactions exist nowadays, e.g. the pragmatic empirical approach of Grimme[26,27], empirically fitted functionals such as M06-2X and most recently special correlation functionals.[28,29]

The most successful density functionals for maingroup chemistry at the moment seem to be the double hybrid functionals which can be regarded as a mixture of hybrid DFT and the wavefunction theory method MP2 where the unoccupied DFT orbitals are used to calculate an additional correlation term by second order perturbation theory.[26,30,31] Their applicability for transition metal chemistry is less clear, however.[32]

Thus there has been progress in development of approximations to the exchange-correlation term and Perdew et al. have argued that the functional forms in use, together form a hierarchy of approximations of increased complexity and accuracy, called Jacob's ladder.[33] On the first rung only the local density is an ingredient of the functional (LDA functionals), on the second rung the gradient of the density is added (GGA functionals), then the kinetic energy density (meta-GGA functionals) on the third rung, occupied-orbital dependent exact exchange comes in on the fourth rung (hybrid functionals) and on the fifth and last rung, unoccupied orbital-dependent correlation is added. While only an empirical classification system, an inverse correlation between mean absolute errors of density functionals for a large maingroup thermochemistry database and the rungs of Jacob's ladder was evident in a recent study.[31,34]

2.2.5 Semi-empirical methods

Semi-empirical methods are another type of electronic structure methods. They are usually based on the Hartree-Fock method and are derived from HF by reducing the number of integrals, resulting in methods that are of much lower computational cost than HF (or DFT). Additionally, they only consider the valence electrons (core electrons are merged with the nucleus to create an effective nuclear charge) and use a minimal basis set (often STOs). The low computational cost of semi-empirical methods comes mainly from reducing the number of two-electron integrals (J_{ij} and K_{ij}), but some one-electron integrals are set to zero as well. Other integrals are then

made into parameters which are fit to either experimental or higher-level theoretical data. The level of integral reduction and how the parameterisation is performed, defines each semi-empirical method.[2]

The more common methods are AM1[35] and PM3[36], based on the modified neglect of diatomic overlap (MNDO) approximation.[37] The PM3 method will be used in Chapter 4 of this work.

The low computational cost of semi-empirical methods make them particularly attractive choices for fast geometry optimisation of large molecules (and molecular dynamics as will be discussed in Chapter 4), but a problem with these methods is that the accuracy for energies and geometries is limited.

2.2.6 Basis sets

Both Hartree-Fock, post-Hartree Fock and DFT methods depend on the use of 1-electron wavefunctions or orbitals to build up the many-electron wavefunction or the electron density.
In a molecule, the self-consistent HF/DFT orbitals can end up having a complicated shape and in order to give the greatest flexibility, it is useful to take linear combinations of many different functions:

$$\phi_i(\mathbf{x}) = \sum_{\mu} c_{\mu i}\varphi_{\mu}(\mathbf{x}) \tag{27}$$

where φ_μ are called basis functions and $c_{\mu i}$ basis set coefficients. In molecular quantum chemistry, atom-centered basis functions are mostly used while in the solid state quantum chemistry/physics community, delocalised planewave basis functions are more common.

The exact wavefunctions of the 1-electron hydrogen atom have served as inspiration for the form of these atom-centered basis functions. The hydrogen atom wavefunctions can be described by products of spherical harmonics and exponential functions, $e^{-\alpha r}$. It turns out, however, that Gaussian functions, $e^{-\alpha r^2}$, are in many ways more suitable to evaluate the integrals of Hartree-Fock and density functional theory. A Gaussian-type orbital (GTO) is usually calculated as:[1]

$$\chi^{GTO}(x,y,z;\alpha,i,j,k) = \left(\frac{2\alpha}{\pi}\right)^{3/4}\left[\frac{(8\alpha)^{i+j+k}i!j!k!}{(2i)!(2j)!(2k)!}\right]^{1/2} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)} \tag{28}$$

using Cartesian coordinates (instead of spherical coordinates), where the integers *i*, *j* and *k* determine the shape of the orbital and α is the exponent. For example: when *i*=*j*=*k*=0 a spherically symmetric s orbital is defined, analogous to the ground state wavefunction of the hydrogen atom.

Gaussian-type orbitals, however, do not resemble the hydrogen wavefunctions as well as their exponential counterparts, usually called Slater-type orbitals (STOs) where the latter have the correct cusp behaviour at the nucleus and the correct radial decay which the GTOs lack. In order to mimic the behaviour of STOs using GTOs, a strategy has been to make linear combinations of a few GTOs (with different exponents of each GTO):

$$\varphi_\tau(\mathbf{x}) = \sum_a d_{a\tau} \chi_a^{GTO}(\mathbf{x}) \tag{29}$$

where φ_τ is called a contracted Gaussian basis function and $d_{a\tau}$ is the contraction coefficient. Usually, only a few GTOs are used in each contracted basis function and importantly the contraction coefficients are fixed and not values to be variationally optimised.

Basis functions intended for a molecular calculation, together form a basis set and the convergence behaviour of basis sets has been much studied over the years. Appropriate exponents, contractions and types of the basis functions are usually pre-optimised for each element and are remarkably transferable. Typically, the lowest energy (core) orbitals change very little from one molecule to the next and remain atom-like. This means that they can be described very efficiently by a low number of basis functions and often a greater degree of contraction than the valence orbitals. This observation lead to the development of the popular split-valence basis set family. A popular split-valence basis set is 6-31G[38,39] which is available for many elements. If the orbital structure of a carbon atom is described by this basis set then that means that the 1s core orbital is described by a single contracted basis function which is a linear combination of 6 GTOs. Each valence orbital on the other hand (2*s*, $2p_x$, $2p_y$, $2p_z$) is described by a linear combination of two contracted basis functions, one of which is described by 3 GTOs and the other by 1 GTOs. In total this adds up to 9 contracted basis functions and 22 GTOs that adequately describe a carbon atom in molecular calculations. The 6-31G basis set is called a double-zeta basis set as the valence orbitals are described by double the amount of contracted basis functions than a minimal description (a minimal basis set uses only 1 basis function for each orbital) for even more flexibility. It turns out that for very accurate calculations, a few contracted basis functions (4-6) for each valence orbital are necessary for accurate post-HF calculations, while triple-zeta basis sets are often sufficient for Hartree-Fock and DFT calculations. However, even a great number of basis functions is not always enough. For an accurate description of the carbon atom, especially when part of a molecule, one cannot rely solely on basis functions of the *s* and *p* type. For additional flexibility of the valence orbitals, polarisation functions are typically added (usually of the next higher angular momentum type). For a carbon atom, a set of *d* polarisation functions (single GTOs) would be added to each p basis function of the 6-31G basis set, resulting in the 6-31G(d) or 6-31G* basis set.[40] Polarisation functions turn out to be crucial to describe molecules well and are almost always included in basis sets. Additionally, diffuse functions are sometimes added to basis sets for a better description of diffuse orbitals, which is usually necessary when calculating anions for example. Low-exponent basis functions (single GTOs) of various types are then added to the basis set. A 6-31G(d) basis set augmented with diffuse functions is named 6-31+G(d).[41]

In this work, we will make use of various basis sets, depending on their applicability for the different molecular systems we calculate. These include:

- The double-zeta polarised basis set 6-31G(d)[38-40] and its augmented version 6-31+G(d).[41]
- The triple-zeta highly polarised and diffuse 6-311+G(3df,3pd) basis set[42] which should be close to the DFT basis set limit.

- The augmented Wachters basis set[43,44] for the first transition-metal row. When combined with the 6-31G(d) basis set on the ligand atoms for transition metal complexes, the combined basis set is referred to as AE1.
- The Turbomole def2 basis set family: def2-SVP, def2-TZVPP, def2-QZVPP which are polarised basis sets of the double-zeta, triple-zeta and quadruple-zeta level, similar in make-up as the split-valence basis sets but available for almost the whole periodic table and show fast convergence when used in DFT calculations.[45]
- The correlation-consistent basis set family of Dunning et al. : cc-pV*n*Z[46,47] where *n*=D, T, Q, 5, 6 and the diffuse versions aug-cc-pV*n*Z.[48] This basis set family has the advantage of systematically approaching the basis set limit and has been designed for use with correlated wavefunction methods where the basis set dependence is much greater than with DFT and HF.

Finally, the use of basis functions is of course an approximation and introduces a basis set error that only goes completely away by using a complete set of basis functions, which is in principle infinite. By carefully choosing appropriate basis functions and studying the convergence of the calculated energy/property with respect to basis functions, the basis set error is made manageable. In certain cases, decontraction of the basis set becomes necessary (the contraction is actually a constraint in a variational calculation), in particular for certain molecular properties where the pre-determined basis set contraction is non-ideal for that particular property. Decontraction turns each GTO into its own basis function and the associated coefficient is now free to variationally optimise during the SCF procedure. This will always lead to a lower energy solution at increased computational cost.

2.3 Geometry optimisations, molecular dynamics and free energy simulations

Geometry optimisations on a 0 K Born-Oppenheimer PES are probably the most common way of using quantum chemical methods to model molecular behaviour. The nuclear motion is ignored and critical points on the surface are located by calculating the atomic forces of the nuclei (differentiation of the total energy with respect to atomic coordinates) and searching for chemically relevant critical points which are either minima or 1st order saddle points. The 0 K potential energy difference between critical points can be regarded as an excellent approximation to conformational energies, reaction energies and transition states of reactions, obtainable from experiments. Corrections for the lack of nuclear motion can be obtained by solving approximations to Eq. 10, e.g. the harmonic approximation, which results in zero-point energy corrections and vibrational frequencies which when combined with approximations for rotational and translational motion from statistical mechanics, yields corrections so that temperature-dependent enthalpies, entropies and free energies can be calculated.

However, there are situations where the dynamic behaviour of the nuclei should be accounted for from the beginning. Molecular dynamics simulations involve solving Newton's equations of motion for the classical nuclei.

Newton's second law relates the force, mass and acceleration of particles: $\mathbf{F}=m\mathbf{a}$
The nuclear forces (and masses) of a molecule can thus be related to velocities and changes of nuclear positions according to:

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} = \frac{\mathbf{F}}{m} = -\frac{dE}{d\mathbf{r}}\frac{1}{m} \tag{30}$$

Solving Eq. 30 enables one to calculate how the nuclear positions change as a function of time. Assuming a molecule with an initial position vector $\mathbf{r}_i$ containing all atomic coordinates, the position vector a timestep later can be given by a Taylor expansion:[2]

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \frac{d\mathbf{r}}{dt}(\Delta t) + \frac{1}{2}\frac{d^2\mathbf{r}}{dt^2}(\Delta t)^2 + \frac{1}{6}\frac{d^3\mathbf{r}}{dt^3}(\Delta t)^3 + \ldots \tag{31}$$

or equivalently:

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 + \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \ldots \tag{32}$$

Different algorithms have been developed to conveniently estimate the position vector a timestep later. One is the Verlet algorithm:[49]

$$\mathbf{r}_{i+1} = (2\mathbf{r}_i - \mathbf{r}_{i-1}) + \mathbf{a}_i(\Delta t)^2 \tag{33}$$

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i}\frac{dE}{d\mathbf{r}_i} \tag{34}$$

The Verlet algorithm is correct to third order (the third term of the Taylor expansion cancels out in the derivation of the algorithm) and easily allows the prediction of the next position using the information of the current and previous positions and the calculated forces. The accuracy of the algorithm is controlled by choosing smaller and smaller timesteps.[2]

A problem with the Verlet algorithm is that velocities are not part of the equations, creating difficulties in using the algorithm to generate ensembles at constant temperature as temperature is usually maintained by modifying the velocities of particles.
In the leapfrog algorithm, however, velocities appear explicitly:[2]

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_{i+\frac{1}{2}}(\Delta t) \tag{35}$$

$$\mathbf{v}_{i+\frac{1}{2}} = \mathbf{v}_{i-\frac{1}{2}} + \mathbf{a}_i(\Delta t) \tag{36}$$

Here the velocity and position updates are out of phase by half a time step. This nevertheless leads to very accurate trajectories and is accurate to third order as the Verlet algorithm but is numerically more stable. The leapfrog method was used in the molecular dynamics simulations in this work.

Any computational method (from quantum mechanics or molecular mechanics) can be used to evaluate the forces in the Verlet and leapfrog algorithms and hence the time-dependent behaviour of molecular systems can be studied under the assumption that the classical behaviour of the nuclei is an accurate enough representation of their real quantum nature. This assumption is generally valid, with the lightest nucleus, hydrogen, being a borderline case.

Molecular dynamics simulations generate trajectories of an ensemble, the most straightforward one to calculate being the NVE ensemble which corresponds to the

particle number, volume and energy being constant during the simulation. As the energy is fixed, although kinetic and potential energy will be interconverted, the temperature can fluctuate considerably but no thermal exchange with the environment is allowed (in fact there is no environment). This corresponds well to the experimental situtation of a molecule in a vacuum. In condensed phase experiments, the temperature is often held constant, and hence the corresponding ensemble to use in simulations is usually either NVT or NPT where in the former the volume and temperature are constant, while in the latter the pressure and temperature are constant.

The NVT ensemble (which is used in this work) is also called the canonical ensemble. In order to make the temperature constant, the simulated system is usually coupled to a heat bath, or thermostat. The Nosé-Hoover method[50-53] is an example of where the thermostat becomes an integral part of the molecular system as an artifical dimensionless particle with a mass and velocity is introduced and equilibrates with the system and exchanges kinetic energy with all real particles of the system. By controlling the equations of motion of this dimensionless particle, any desired temperature can be set while preserving all the features of the Verlet/leapfrog algorithm and the NVT ensemble.

The free energy is the thermodynamical quantity directly related to the equilibrium constant and is related to the probability that a system will be in a particular state. The ability to determine free energy differences from molecular dynamics simulations is highly useful as they can then be directly related to the thermodynamical data from experimental measurements.

The Helmholtz free energy of a system in the NVT ensemble is given by:

$$A = -k_B T ln Q \tag{37}$$

where *Q* is the partition function of the system which is expressed classically as a double integral over all energy states which are functions of spatial and momentum coordinates:[1]

$$Q = \int \int e^{-E(\mathbf{p},\mathbf{r})/k_B T} d\mathbf{p} d\mathbf{r} \tag{38}$$

It is very hard to determine free energy differences of a system by direct calculation of free energies by the equations above due to the large number of degrees of freedom that need to be sampled. However, useful methods to calculate free energy differences from molecular dynamics simulations nevertheless exist. The free energy perturbation expression:

$$\Delta A = \langle A \rangle_b - \langle A \rangle_a = k_B T \, ln \langle e^{(E_b - E_a)/k_B T} \rangle_a \tag{39}$$

allows direct calculation of the free energy difference between state *a* and state *b* where all the ensemble averaging is based on state *a*.[1] As long as state *a* and state *b* are sufficiently similar (i.e. the perturbation is small), this is a very useful approximation. It can additionally be expanded so that the free energy difference between *a* and *b* is calculated by dividing up the path connecting *a* and *b* into intermediate states, calculating the energy difference between each intermediate state and add everything up in the end.

Another very useful method is the thermodynamic integration technique where the free energy difference can be determined by integration of the ensemble average of the potential energy gradient w.r.t. fixed values of the reaction coordinate. For certain reaction coordinates this becomes equal to integrating over the ensemble average of the force of the constrained reaction coordinate:[54,55]

$$\Delta A^{a\to b} = \int_{\xi a}^{\xi b} \left\langle \frac{dE}{d\xi} \right\rangle_{\xi} d\xi = -\int_{\xi a}^{\xi b} \langle f(\xi)\rangle_{\xi} d\xi \tag{40}$$

A reaction coordinate ξ, connecting the two states is thus identified and constrained. A number of simulations are then run, constraining different values of the reaction coordinate while sampling the force of the constraint $f(\xi)$. Integrating over the average constrained force for each simulation (the more simulations the better) results in the free energy difference between state *a* and *b*. In this work, the reaction coordinate will be a geometric variable (although it can be any variable shared by state *a* and *b*).

2.4 Transition state theory

Transition state theory[56-58] is currently the most approachable rate theory of chemical reactions that can relate the results of quantum mechanics calculated potential energy surfaces to reaction rates. The reaction rate is governed by the free energy difference between the reactant state and a transition state (activation free energy) as the simple Eyring equation shows:[59]

$$k = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \tag{41}$$

Transition state theory assumes that the reactant state is in equilibrium with the transition state and that the system only crosses the transition state once (no re-crossing allowed) as well as the reaction occuring on a single Born-Oppenheimer PES and that the nuclei are described classically.

The transition state for a chemical reaction can to a good approximation be taken as the saddle point on a calculated potential energy surface and the potential energy barrier is usually a good approximation to the free energy barrier (when entropic effects are small). Predicted activation energies that are less than 25 kcal/mol, can as a rule of thumb be assumed to occur spontaneously at room temperature according to transition state theory. Limitations of transition state theory are mostly significant at low and high temperature. At low temperature, an accurate representation of zero-point motion may be necessary and quantum mechanical tunneling through the barrier may start to affect the rate. Extensions to transition state theory are a topic of active research.

2.5 Molecular properties

Once the solution of the Schrödinger equation has been approximated by any of the methods in Chapter 2.2, one can turn to molecular property calculations. One way of expressing molecular properties is in the language of analytical derivatives. We wish to see how the energy changes when a small perturbation is applied and so we

perform a Taylor expansion of the energy with respect to the perturbation. This involves calculating derivatives of the energy, sometimes resulting in complicated expressions for each of the terms of the energy functional used to solve the Schrödinger equation, especially if second derivatives or higher are required.[8] The following subchapters discuss the electric field gradient (a first order property, where only the first derivative of the energy is required) and the chemical shift (a second order property, where the second derivative of the energy must be calculated).

2.5.1 Electric field gradients (EFG)

The interaction between a quadrupolar nucleus (a nucleus with spin $I > ½$) and its environment, can be written as an operator:[60]

$$H_Q = e^2 \sum_{A,i} Q^{(A)} \mathbf{I}^{(A)} \mathbf{F}^{(A)}(i) \mathbf{I}^{(A)} \tag{42}$$

where e is the elementary charge, $Q^{(A)}$ is the electric quadrupole moment of nucleus A, $\mathbf{I}^{(A)}$ is the spin-orbit operator of nucleus A and the $\mathbf{F}^{(A)}$ is the field gradient operator of nucleus A. The matrix element of the electronic field-gradient operator is defined as:[60]

$$F_{\mu\nu}^{(A)}(i) = \frac{r_{iA}^2 \delta_{\mu\nu} - 3 r_{iA;\mu} r_{iA;\nu}}{r_{iA}^5} \tag{43}$$

where μ and ν are the Cartesian components (μ,ν ∈{x,y,z}), $\delta_{\mu\nu}$ is the delta function and r_{iA} is the distance between electron i and nucleus A.

The electronic contribution to the EFG tensor V is directly calculated using the available density matrix from a converged SCF calculation :[8]

$$V_{\mu\nu,ele}^{(A)} = \sum_{\kappa,\tau} P_{\kappa\tau} \langle \varphi_\kappa | r_A^{-5} (r_A^2 \delta_{\mu\nu} - 3 r_{A;\mu} r_{A;\nu}) | \varphi_\tau \rangle \tag{44}$$

where $V^{(A)}_{\mu\nu,\,ele}$ is the second-rank EFG tensor, $P_{\kappa\tau}$ is the density matrix (the matrix of all optimised basis set coefficients from the SCF calculation) and φ are the basis functions.
The nuclear contribution is then calculated by:[61]

$$V_{\mu\nu,nuc}^{(A)} = \sum_{Y \neq A} Z_Y \frac{r_A^2 \delta_{\mu\nu} - 3 r_{A;\mu} r_{A;\nu}}{r_A^5} \tag{45}$$

a classical equation as the nuclei are approximated as static point charges. The total EFG tensor is finally $V = V^{(A)}_{\mu\nu,\,ele} + V^{(A)}_{\mu\nu,\,nuc}$.
The full EFG tensor, V, can be represented as a matrix in the Cartesian coordinate system:[62]

$$V = \begin{vmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{yx} & V_{yy} & V_{yz} \\ V_{zx} & V_{zy} & V_{zz} \end{vmatrix} \tag{46}$$

The EFG tensor is symmetric ($V_{xy} = V_{yx}$ and so on) and can additionally be diagonalised (i.e expressing it in the principal axis system), which makes the off-diagonal elements vanish and results in the three principal components (or eigenvalues), here called V_{xx}, V_{yy}, V_{zz} and three eigenvectors, describing the orientation of the tensor in space. The EFG is non-zero only when the charge distribution around the nucleus violates cubic symmetry. The largest of the eigenvalues (here defined to be V_{zz}) is then used to calculate the nuclear quadrupole coupling constant (NQCC) :

$$C_Q = \frac{eQV_{zz}}{h} \tag{47}$$

and the asymmetry parameter is defined as follows:

$$\eta_Q = \frac{V_{xx} - V_{yy}}{V_{zz}} \tag{48}$$

where $V_{zz} \geq V_{xx} \geq V_{yy}$.
These two parameters are the only ones (other equations involving the product of Q and V_{zz} are also sometimes used) needed to fully describe the quadrupole interaction as it occurs in spectra due to the trace of the tensor being zero ($V_{zz} + V_{yy} + V_{xx} = 0$). The NQCC and the asymmetry parameter are usually refined during line shape analysis of solid-state NMR spectra as well as gas-phase microwave (MW) spectra.

The EFG tensor is usually calculated in atomic units (au) by quantum chemistry codes. We note that different sign conventions are in use, e.g. the one used in the Gaussian 03[63]/09[64] programs requires multiplication by -1 in order to get the correct sign for the NQCC for comparison with experimental data (when the sign of the NQCC is available).[62]

In our work, Gaussian-style V_{zz} values are always reported but C_Q values from calculations have always been converted to the correct sign (the experimental sign). The following equation is used for convenience:

$$C_Q = kQV_{ZZ} \tag{49}$$

where V_{zz} is in au, C_Q is in MHz, Q is in fm^2 and k = -2.34964781 au^{-1} Bohr^2 fm^{-2} MHz, is a constant that takes care of unit conversion, includes both e and h and takes care of the correct sign use for C_Q.

2.5.2 NMR chemical shifts

NMR-active nuclei are those with nuclear spin $I > 0$ and have $2I+1$ nuclear energy levels, all degenerate under no applied field. When a field is applied, the degeneracy is lifted, opening up the possibility of performing spectroscopy by exciting the nucleus from one energy level to the other.
The resonance frequency is the frequency of electromagnetic radiation that will excite the nucleus to another energy level and is defined as:

$$\nu = \frac{\gamma B_0}{2\pi}(1 - \sigma) \tag{50}$$

where ν is the resonance frequency, γ the gyromagnetic ratio of the NMR nucleus (ratio of the magnetic dipole moment and its angular momentum) for each nuclide, B_0 the applied magnetic field and σ the shielding. The shielding results in the effective magnetic field at the nucleus to be different than the applied field. In NMR experiments, the resonance frequency is thus dependent on both the applied field and the shielding. Since the interest lies in using the magnetic resonance as a probe to chemical properties, the shift in resonance frequency is reported relative to a reference compound resulting in field-independent chemical shifts, δ.

The experimental chemical shift, δ, is therefore directly related to the shielding:

$$\delta = \frac{\nu_{sample} - \nu_{ref}}{\nu_{ref}} \times 10^6 = \frac{\sigma_{ref} - \sigma_{sample}}{1 - \sigma_{ref}} \approx \sigma_{ref} - \sigma_{sample} \tag{51}$$

The last approximation seen in the Eq. 51 can be made because σ_{ref} is usually very small compared to unity.

From a non-relativistic quantum mechanical perspective, the shielding, or the magnetic shielding tensor σ^N_{ij} of a nucleus N, is the mixed second derivative of the total molecular energy E with respect to the magnetic moment μ^N and the external magnetic field B:[65]

$$\sigma_{ij}^N = \left(\frac{\delta^2 E}{\delta B_i \delta \mu_j^N}\right)_{\mathbf{B},\mu^N=0} \tag{52}$$

Computationally, the shielding is usually calculated as a sum of diamagnetic and paramagnetic contributions:[8]

$$\begin{aligned} \sigma_{ij} &= \sigma_{ij}^{(d)} + \sigma_{ij}^{(p)} \\ \sigma_{ij}^{(d)} &= \frac{\alpha^2}{2} \sum_{\mu\nu} P_{\mu\nu} \langle \varphi_\mu | r_A^{-3} [\mathbf{r}_A \mathbf{r} - \mathbf{r}_{A,i} \mathbf{r}_j] | \varphi_\nu \rangle \\ \sigma_{ij}^{(p)} &= \frac{\alpha^2}{2} \sum_{\mu\nu} \frac{\delta P_{\mu\nu}}{\delta B_i} \langle \varphi_\mu | (-i\nabla \times (\mathbf{r} - \mathbf{R}_A))_j r_A^{-3} | \varphi_\nu \rangle \end{aligned} \tag{53}$$

It is the paramagnetic part $\sigma^{(p)}_{ij}$ that is slightly troublesome to compute as it involves the derivative of the density matrix. This necessitates the use of the coupled-perturbed SCF equations making chemical shift calculations more expensive than most other properties.[8]

A further complication arises because the magnetic operators involved depend on the origin of the coordinate system leading to artificial origin-dependent results. This is called the gauge problem. If a complete basis set is used, the results will be independent of the chosen origin, which is of course never the case in practice. Different ways of dealing with this gauge-dependency have been proposed, the most satisfactory solution probably being the use of magnetic-field dependent basis sets, the so called gauge-including atomic orbitals method (GIAO).[66] With this method, the basis functions are made magnetic-field dependent by multiplying them by a phase factor referring to the position of the nucleus:[8]

$$\tilde{\varphi}_\mu(\chi, \mathbf{B}) = \varphi_\mu^A exp(i(\alpha/2)(\mathbf{B} \times \mathbf{R}_A)\mathbf{r}) \tag{54}$$

The shielding is, due to its mixed second derivative form, an unsymmetrical tensor with 3 × 3 different elements.

$$\sigma = \begin{vmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{vmatrix} \tag{55}$$

The matrix can be split up into a symmetric part and an antisymmetric part. The antisymmetric part typically does not influence NMR spectra and is thus not of interest. The symmetric part of the tensor, $\sigma^{sym} = ½ (\sigma + \sigma^{t})$, where σ^{t} is the transpose of σ, can be diagonalised if expressed in the principal axis system, giving three orthogonal eigenvectors and three eigenvalues, σ_{11}, σ_{22} and σ_{33} that describe the position and shape of the NMR signal (ignoring *J* coupling, quadrupole coupling and other effects). The eigenvectors describe the orientation of the tensor components in space.[62]
The average of the eigenvalues is the isotropic shielding, $\sigma_{iso} = ⅓ (\sigma_{11} + \sigma_{22} + \sigma_{33})$.

The isotropic chemical shift, δ_{iso}, of a nucleus in a molecule is evaluated by taking the difference in the isotropic shielding between it and the nucleus in an accepted reference molecule (usually the one that defines the origin of the chemical shift scale).

$$\delta_{iso} = \sigma_{ref} - \sigma_{iso} \tag{56}$$

Computational chemical shifts are hence evaluated in an analogous way to NMR experiments.

The principal components of the chemical shift tensor, that can be obtained from a solid-state NMR spectrum, can be represented in several different ways. In this work we follow the Haeberlen-Mehring-Spiess convention[67-69] as shown in Eqs. 57-60 and Figure 1.

$$\delta_{ii} = \sigma_{iso,(ref)} - \sigma_{ii} (i = 1, 2, 3) \tag{57}$$

$$|\delta_{zz} - \delta_{iso}| \geq |\delta_{xx} - \delta_{iso}| \geq |\delta_{yy} - \delta_{iso}| \tag{58}$$

$$\delta_{\sigma} = \delta_{zz} - \delta_{iso} \tag{59}$$

$$\eta_{\sigma} = \frac{\delta_{yy} - \delta_{xx}}{\delta_{\sigma}} \tag{60}$$

where δ_{σ} is called the reduced anisotropy and η_{σ} the asymmetry parameter.

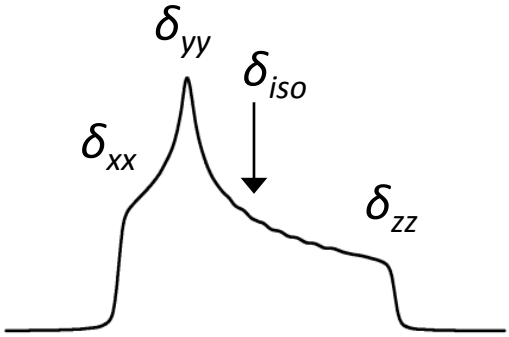


Figure 1 A static solid-state NMR spectrum showing a specific assignment of the principal components according to the Haeberlen-Mehring-Spiess convention. Under magic angle spinning (MAS) conditions this spectrum would turn into a spinning-sideband pattern with an envelope similar to this figure.

Using this convention, the isotropic chemical shift gives the position of the NMR signal on the chemical shift scale (a relative shielding) while the reduced anisotropy and the asymmetry parameter describe the shape of the signal relative to the isotropic chemical shift.

Finally, the spin-spin coupling constant is a second order property like the chemical shift and is defined as the second derivative of the energy w.r.t. the magnetic moments of the nuclei. It involves the separate computation of four different terms that together contribute to the coupling pattern that often shows up in NMR spectra.[70] These are the Fermi contact term (K^{FC}), the spin dipole term (K^{SD}), the paramagnetic spin-orbit term (K^{PSO}) and the diamagnetic spin orbit term (K^{DSO}), as originally deduced by Ramsey.[71]

Computation of NMR properties usually involves single-point calculations on previously optimised geometries using any quantum mechanics based method for which the necessary molecular property integrals have been coded. Normal basis sets can often be used, although it has been found that the addition of tight (large exponent) *p* functions for shielding constants[72] and tight *s*, *p*, *d* and *f* functions for spin-spin couplings[73] can sometimes improve results.

2.6 Partial atomic charges

Partial atomic charge is a much used concept in chemistry that gives at least a useful qualitative description of the charge distribution in a molecule.
Atomic charge, however, is a rather ill-defined property. The problem is defining where one atom ends and another begins in a molecule when dividing up the electron density.
Additionally, partioning the electronic density into a simple monopole distribution will always be an approximation that is most easily improved upon by calculating higher order multipoles instead of just atomic charges. Nevertheless, atomic charges remain very useful: they are used to describe the Coulomb component in force fields for example and can be used to define a simple embedding potential in electrostatic embedding QM/MM.

One of the simplest ways of calculating partial atomic charges from the result of an electronic structure calculation is by calculating Mulliken charges. Here the electrons are divided up so that they belong to an atom according to how the atom-centered

basis functions contribute to the molecular orbitals. One of the problems with Mulliken analysis, however, is that when it is performed using a non-orthogonal basis set, artifacts are created.[1] Additionally, in Mulliken analysis, shared electrons are divided up equally between the atoms on which the basis functions are, meaning that Mulliken analysis ignores the fact that some atoms may have very different electronegativities. The basis set sensitivity of Mulliken charges is also a problem.[74]

The Natural Population Analysis (NPA) scheme by Weinhold et al.[75] is a significant improvement upon Mulliken analysis. Here the basis set is first made orthogonal by a complex orthogonalisation scheme, before a Mulliken-like division of electrons is performed. The NPA scheme thus removes the artifacts of using a non-orthogonal basis set and the charges are much better behaved with respect to increasing basis set size.

ESP charges belong to another type of partial atomic charges. The electrostatic potential (ESP) is calculated on points on a vdW surface around the molecule and atomic charges are fitted so that they reproduce the ESP and the total charge.[76,77]

None of these charge models are perfect but are useful for different purposes. In this work we have mostly used NPA charges but have also compared them with Mulliken and ESP charges in some chapters.

2.7 Continuum solvation models

Continuum solvation models account for solvent effects implicitly by describing the solute by a charge distribution (often quantum mechanical) in a (usually molecule-shaped) cavity in a dielectric medium that describes the solvent, and the interaction between solute and solvent. The dielectric medium is defined essentially by the dielectric constant (or sometimes the relative permittivity (a scalar function of position)), which can be understood as the polarity of the solvent (water being polar has a high dielectric constant, while hexane being nonpolar has a low dielectric constant). The charge distribution of the solute can polarise the dielectric medium and the dielectric medium can polarise the solute and the total electrostatic potential due to the mutual polarisation of solute and solvent is obtained by solving the nonhomogenous Poisson electrostatics equation: [1,78]

$$\nabla \cdot (\epsilon \nabla \Phi) = -4\pi\rho_f \tag{61}$$

where ∇ is the Nabla operator, ε the dielectric constant, Φ the total electrostatic potential and ρ_f the solute charge distribution. If solved in a vacuum (ε =1) with ρ_f approximated as a positive point charge (a crude approximation of a cation), the Poisson equation leads to the simple Coulomb potential of a point charge in a vacuum. If ε is something else and ρ_f a more complicated charge distribution, the total potential that is obtained is the simple vacuum Coulomb potential plus a reaction field:

$$\Phi = \phi^{vac} + \phi^{rxn} \tag{62}$$

Continuum solvation models that use a quantum mechanics based description of the solute ρ_f will only be mentioned in this work (others use a classical charge distribution, e.g. Generalized Born and do not start from the Poisson equation).[1]

The (electrostatic) free energy of solvation is then obtained by solving the following equation:

$$\Delta G_{elstat} = \langle \psi | H^0 - \frac{e}{2}\phi^{rxn} | \psi \rangle + \frac{e}{2}\sum_k Z_k \phi_k^{rxn} - \langle \psi^0 | H^0 | \psi^0 \rangle \tag{63}$$

where ψ is the polarised wavefunction, ψ^0 the unpolarised wavefunction, H^0 the normal Hamiltonian, e the elementary charge and ϕ^{rxn} the reaction field. Solving the Schrödinger equation (WFT,DFT) in the presence of the reaction field leads to a new electrostatic potential, that leads to a new reaction field; these equations must thus be iterated until self-consistence. A well known continuum solvation model is the polarisable continuum model (PCM)[3] where the Poisson equation is solved for a set of charges (approximating the reaction field potential) that sit on the cavity surface (boundary element method) and polarise the quantum charge distribution. The cavity is defined by combining atom-centered spheres that together define a surface.

Many other solvation models are variants of PCM and differ mainly in the way the cavity is defined or how non-electrostatic effects are taken into account. In the IPCM[79] method for example, the cavity surface is defined by a constant value of the gas-phase electron density and in the SCIPCM method[79] the same surface is iterated self-consistently. In the recent SMD solvation model,[78] the standard PCM model is used (or rather the modern integral-equation-formalism) but using a special set of atomic radii that define the atomic spheres, cavity and surface and an additional component that accounts for non-electrostatic effects.

Drawbacks of continuum solvation models include the approximation of describing a large complicated chemical system (ideally described explicitly by quantum mechanics) by a simple polarised dielectric medium as well as ambiguities involving the size and shape of the cavity, approximations of the non-electrostatic effects (cavitation, dispersion, exchange-repulsion, charge-transfer etc.) if included at all, and many others.

2.8 QM/MM methods

The idea behind quantum mechanics / molecular mechanics methods (QM/MM) is to retain the atomic degrees of freedom of the real molecular system one wants to model, by replacing most of the full quantum mechanical description with molecular mechanics (an approximation to quantum mechanics). The quantum mechanical part (usually in the center of the system) thus describes the main region of interest while molecular mechanics acts as an approximation to the environment surrounding the quantum mechanical part.

Almost all QM/MM methods can be described by the additive equation:[80]

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}}(\mathrm{I}) + E_{\mathrm{MM}}(\mathrm{II}) + E_{\mathrm{QM\text{-}MM}}(\mathrm{I}, \mathrm{II}) \tag{64}$$

Where $E_{QM/MM}$ is the total energy of the system, E_{QM}(I) is the QM energy of the QM region (I), E_{MM} is the MM energy of the MM region (II) and $E_{QM\text{-}MM}$(I , II) is the QM-MM interaction term between the two regions.

The E_{QM}(I) term is simply a normal quantum mechanics calculated energy using any suitable method. The only prerequisite is that the QM method must be able to perform energy and gradient calculations in the presence of external point charges when using the electrostatic embedding scheme (see later).
The E_{MM} energy expression consists of both bonded (bonds, angles and dihedrals) and nonbonded (Lennard-Jones potential and the Coulomb potential of point charges) terms:

$$E_{MM} = \sum_{bonds} k_d(d-d_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} k_\phi[1+cos(n\phi+\delta)] + \sum_{nonbonded} \epsilon_{AB}[(\frac{\sigma_{AB}}{r_{AB}})^{12} - (\frac{\sigma_{AB}}{r_{AB}})^{6})] + \frac{1}{4\pi\epsilon_0}\frac{q_A q_B}{r_{AB}} \quad (65)$$

where the constants k_X, d_0, θ_0, δ, ε, σ, q_A, q_B are fitted so that the force field reproduces conformational energies, interaction energies etc. of molecular systems for which the force field is intended. Several parameterised force fields are available to describe proteins, for example the CHARMM force field[81,82] which is used in this work.

It is the $E_{QM\text{-}MM}$(I , II) term that essentially distinguishes different QM/MM methods from one another (apart from the use of different force fields) as it takes care of the coupling terms (bonded, van der Waals (vdW) and electrostatic interactions) between the different regions:

$$E_{QM\text{-}MM}(I , II) = E^{b}_{QM\text{-}MM} + E^{vdW}_{QM\text{-}MM} + E^{el}_{QM\text{-}MM} \quad (66)$$

The bonded and vdW terms of $E_{QM\text{-}MM}$(I,II) are described at the MM level. The vdW interaction is usually described by a Lennard-Jones potential, necessitating Lennard-Jones parameters for all QM atoms. In theory, every atom of the QM region interacts with every atom of the MM region, but in practice only the atoms closest to the boundary contribute significantly.

The electrostatic QM-MM interaction, which is the most important term, can be described by several different methods, the models can be classified by their increasing complexity, as mechanical embedding, electrostatic embedding and polarised embedding. In the mechanical embedding scheme the electrostatic interations between QM atoms and MM atoms is handled at the MM level where the MM charge model (usually rigid atomic point charges) is used to calculate the interaction between QM and MM atoms where point charges for all QM atoms must be defined. Mechanical embedding is most often used in a subtractive QM/MM approach where the E_{MM} term is calculated for the whole system (I+II) right from the beginning. The main problem with mechanical embedding is that the QM-based electron density is simplified into a simple model of point charges which is additionally not influenced directly by the charges of the MM region (usually).

The electrostatic embedding scheme is a natural extension to mechanical embedding. The electrostatic interaction is simply calculated at the same time as the E_{QM}(I) term by incorporating the point charges of the MM region as an electron point-charge term

and a nucleus point-charge into the Hamiltonian of the QM calculation of the QM region. The nucleus point-charge term is straightforward and the electron point-charge term is simply calculated in the same way as the electron-nucleus term is calculated. The quantum mechanical electron density is thus influenced by the surroundings directly and can adapt to any changes occurring in the MM region.

A polarised embedding scheme takes electrostatic embedding one step further by using a MM charge model that can be polarised as well (by the QM electron density), sometimes involving completely mutual polarisation in which case a self-consistent cycle is required at each QM/MM energy evaluation. Several different polarised embedding approaches have been suggested: polarised point dipoles scheme where assigned atomic polarisabilities induce point dipoles, Drude oscillator/shell/charge-on-spring models where an additional opposite-sign point charge is attached to the MM atom connected by a harmonic spring creating a dipole, fluctuating charges and more.[80]

Electrostatic embedding and polarised embedding QM/MM can also be looked at in the general context of effective embedding potentials. The QM system is embedded by an embedding potential, just like the effective potentials in the single electron HF/Kohn-Sham equations. In electrostatic/polarised embedding QM/MM, the embedding potential is (usually) a simple Coulomb point charge potential (the point charge approximating the multi-particle atom), an approximation to the real smeared out Coulomb multi-particle term of the surroundings. Continuum solvation models can also be seen as embedding potentials. A higher-order multipole potential goes a step further towards the real potential but still neglects exchange-repulsion effects that can in principle be included as well in an embedding potential. The most complicated embedding potentials are those involving whole electron densities, usually frozen, as in the frozen density embedding methods.[83-86]

A specific problem remains that concerns the division of the system into QM and MM regions. Covalent bonds often end up being cut by the QM-MM boundary, especially for systems such as proteins. As the dangling bond of the QM region needs to be saturated, one of the simplest solutions is to include link atoms (usually H) as part of the QM region in the QM energy calculation. Since the added link atoms constitute an additional energy term, it should be corrected for, in practice though, this is rarely done. However, since the link atom will be very close to the first MM atom at the boundary, a strong artificial electrostatic interaction will be present. One approach to solve this problem is called charge-shifting which removes the charge from the MM atom and divides it among the nearest bonded MM atoms.[87,88]

Other methods to deal with the QM-MM boundary are boundary-atom schemes where the MM atom involved in the cut is transformed into a special boundary atom which mimics both the QM interaction and the MM interaction, and localised-orbital schemes where hybrid orbitals are placed at the boundary (saturating the dangling bond) and frozen so that they are not part of the SCF iterations.[80]

QM/MM calculations can nowadays be performed using many different programs, often these are quantum chemistry programs that have incorporated some molecular mechanics functionality or the other way around.
Chemshell, on the other hand, is a computational chemistry program which is neither.[88,89] It uses a modular approach that links together different external

programs through a programmable interface based on the Tool Command Language (Tcl). The main speciality of the program is in hybrid QM/MM simulations where in principle any quantum chemistry program can be linked to any molecular mechanics program to perform QM/MM calculations where the Chemshell program is responsible for the geometry optimisations/molecular dynamics simulations, the QM/MM coupling and the data management. The program allows one to do geometry optimisation and molecular dynamics simulations on small to large chemical systems using a variety of different quantum chemistry programs and molecular mechanics programs through pre-programmed interfaces.

Chapter 3. Modelling environmental effects on chemical shift and EFG properties: From molecular crystals to metalloenzymes

3.1 Introduction: experimental and computational NMR spectroscopy

Nuclear magnetic resonance, NMR, is one of the most important analytic tools to investigate molecular structure as it gives direct information about the chemical environment by measuring a difference in shielding as a magnetic field is applied around an NMR active nucleus. It is also a non-destructive technique as the electromagnetic radiation is on the radiowave scale.

Standard NMR spectroscopy is most often applied to well behaved I=½ nuclides like ^{1}H. The ^{1}H nuclide has very favourable magnetic properties (e.g. a large gyromagnetic ratio) and high abundance, making it an excellent probe to hydrogen-containing molecules. With better spectrometers and the Fourier-Transform technique, it has become possible to measure other I=½ nuclides with lower abundance like ^{13}C and ^{15}N. However, I>½ nuclides have always been troublesome due to quadrupolar broadening effects on the spectra, resulting from the quadrupole interaction.

Transition metal NMR spectroscopy has the advantage of being a direct probe to the chemistry of a metal complex and transition metal nuclei have a very large chemical shift range (many thousands of ppm) and thus a much larger sensitivity to structural changes than the ^{1}H nucleus. There are problems, however, associated with small magnetic moments, small to large nuclear quadrupole moments leading to broadness of peaks (especially in solution), a range of different relaxation times, low abundance and a generally lesser sensitivity than ^{1}H NMR.[90] But with spectrometers becoming more powerful and new detection methods being constantly developed, NMR spectroscopy of transition metal nuclides is gradually becoming a very useful and approachable spectroscopic method.

There has been much progress in solid-state NMR spectroscopy techniques over the years and solid-state NMR for quadrupolar nuclides is much more approachable than solution NMR despite the fast quadrupolar relaxation that can lead to very large peak broadening effects.[91] A static solid-state NMR spectrum is typically a broad featureless spectrum (see Figure 1) that can be hard to analyse as the nucleus in a condensed phase experiences many interactions. By spinning the sample at a specific angle (the magic angle: 54.74°) the main three orientation dependent interactions (dipolar coupling, chemical shift anisotropy and quadrupolar coupling) can be averaged (out, in the case of dipolar coupling), resulting in a much clearer spectrum with a spinning sideband pattern (multiples of the spinning speed). This is called magic angle spinning (MAS). In solid-state NMR spectroscopy, the orientation of the molecule will influence the spectrum as the medium itself is ordered (in contrast to solution NMR) and the chemical shift and quadrupole coupling are orientation dependent. This orientation dependency or anisotropy is described by two molecular property tensors, together essentially defining the complete spectrum that is obtained. These are the chemical shift tensor and the nuclear quadrupole coupling tensor, the latter of which is determined by the EFG (electric field gradient) tensor.

NMR spectroscopy can be an especially important technique for systems containing so called spectroscopically silent atoms, i.e. systems that show featureless UV/VIS

spectra and are diamagnetic. These include biologically prevalent metal atoms such as magnesium, calcium, sodium and zinc. The diamagnetism prevents electron paramagnetic resonance (EPR) spectroscopy from being applied which is one of the most valuable spectroscopic techniques in transition metal chemistry. There is in fact great potential for NMR spectroscopy in the field of bioinorganic and medicinal inorganic chemistry as highlighted in a recent review[92], since metal ions act as important catalytic cofactors of enzymes, as structure-stabilising ions of large molecules, and even as enzyme inhibitors in the case of metallodrugs. In fact one third of all enzymes contain tightly bound metal ions.[93] With better instrumentation and detection methods, transition metal NMR spectroscopy in bioinorganic chemistry can yield crucial structural or even dynamical information, e.g. probing a single metal ion inside a cofactor of a large enzyme becomes possible. There are already studies where this has been achieved: recent examples are ^{51}V NMR of the vanadate cofactor of vanadium chloroperoxidase,[94] ^{67}Zn NMR of human carbonic anhydrase,[95] ^{25}Mg NMR of DNA repair protein APE1[96] and ^{65}Cu NMR of the blue copper protein azurin.[97]

NMR properties of metal nuclei are not as well known as others and this is where first principles calculations can be very useful. Traditional ^{13}C and ^{1}H NMR have become such useful spectroscopic techniques because the shielding environment is flexible but also because much data has accumulated over the years and thus many empirical models exist to help interpret spectra. As NMR spectroscopy of transition metal nuclides is not as popular yet, empirical models are not as available. This becomes the strength of quantum chemical calculations, to predict and explain NMR properties of molecules from first principles, in the not yet mature area of transition metal NMR spectroscopy.

Density functional theory is usually the electronic structure method of choice for NMR computations as second-order properties such as chemical shifts are very expensive to calculate with most higher order WFT methods[65] and low order WFT methods based on Møller-Plesset theory usually do not do so well for transition metal compounds.[98-100] Hybrid density functionals have been found to have an advantage over GGA functionals for the chemical shifts of some transition metal nuclides although currently no single functional is suitable for all nuclides.[101] There have been studies that go beyond static calculations and take both zero-point geometric and thermal averaging geometric effects into account by DFT-based molecular dynamics. Overall it has been found that such effects tend to be small (with exceptions) and are often overshadowed by the error introduced by the DFT approximation.[101] Solvent effects on transition metal chemical shifts have also been studied for vanadium, manganese, iron and cobalt complexes [102-105]. Overall, the solvent effects tend to be small with some exceptions and the effect is usually of a geometric nature. Chapter 3.2 discusses benchmarking of DFT functionals in more detail for ^{51}V chemical shift calculations. Much less is known about computations of quadrupole coupling parameters that influence solid-state NMR spectra of quadrupolar nuclei and this will be the subject of Chapter 3.3.

The solid state presents new challenges for molecular property calculations compared to the gas and liquid states. Calculations of molecular properties that are to be compared with solid-state experimental data, are often performed in the gas-phase, and a common theme in many such studies in the literature is the question of whether quantum mechanics optimised structures or X-ray structures should be used.[106-110] The X-ray structure represents a thermally averaged crystal-packed

structure, but it is taken out of its solid environment and may not always be accurate. In particular, static or dynamic disorder of the crystal, if gone undetected, can introduce severe artifacts on individual geometric parameters, which can have detrimental effects on subsequent molecular property evaluations. Quantum mechanical geometry optimisations in the gas-phase, on the other hand, afford equilibrium structures lacking effects from thermal motion or the environment.

Two QM/MM-based approaches, called the embedded ion method (EIM)[111] and extended embedded ion method (EEIM), have recently been successfully used to calculate NMR properties of molecular crystals and ionic solids[112-115] using self-consistent point charges and Ewald lattice sums. A QM/MM approach has also recently been used to model environmental effects on charge density distributions in the methyl lithium crystal.[116] These embedded molecular property calculations use either experimental X-ray geometries or in some cases optimised hydrogen positions. Typically, QM/MM studies require a full-fledged force field (i.e. well-parameterised bonded and non- bonded terms of the whole molecule) for the MM region in order to perform geometry optimisations. This is not a problem when modeling proteins, for instance, for which many bespoke, well-parameterised force fields exist. On the other hand, for many inorganic or organometallic molecules, and especially transition metal complexes, force field parameters often are unavailable or of limited transferability.

An elegant way to model the solid state comes from periodic DFT approaches where the periodicity of the crystal is directly taken into account in calculations of the unit-cell. In recent years, NMR calculations with periodic DFT approaches have become possible.[117-121] These methods are very promising for modelling NMR properties in any kind of solid as they capture the environmental effects of the solid directly by design. Disadvantages of these approaches are that the core electrons are usually necessarily approximated by pseudopotentials which can affect NMR properties severely. The projector augmented wave (PAW) approach, originally by Blöchl[122] has been used to recover the proper nodal structure of the valence MOs in the core region in pseudopotential planewave NMR calculations with the gauge including projector augmented wave (GIPAW) method.[117,118] It is not always clear how sensitive the results are on the pseudopotentials that are used in the GIPAW method. Additionally, periodic DFT approaches are often limited to GGA DFT methods due to the difficult and expensive implementation of Hartree-Fock exchange in periodic DFT codes although efforts towards more efficient Hartree-Fock exchange have become more successful.[123,124] Finally, periodic DFT calculations may not be the best option if large molecular systems with large unit cells or little symmetry are to be computed.

We are interested in accounting for environmental effects on NMR properties, especially in the area of solid-state NMR spectroscopy where environmental effects are potentially more prevalent than in solution NMR spectroscopy. We distinguish between two types of environmental effects:

1. Environmental effects that affect the geometry of the molecule compared to the geometry of the molecule in a vacuum which in turn affects the property.
2. Environmental effects that directly affects the property of the molecule by perturbing the electron wavefunction/density.

Experimentally, it is very difficult to distinguish between the two, but computationally this is rather straightforward and quite important. The former type

is expected to be more important, NMR properties are known for being sensitive to geometrical effects and any environmental effect that affects the geometry will thus affect the NMR property.
The second type should be less important as NMR properties are usually thought of as short-range properties, being mainly influenced by their nearest environment (atoms that are bound covalently or by strong noncovalent interactions like hydrogen bonds) but this may be highly system-dependent and not obvious from the outset. How environmental effects affect the quadrupole interaction is not really known.

Environmental effects on NMR properties will in this work mainly be explored by cluster and QM/MM (embedded cluster) approaches. This entails approximating the solid as a truncated cluster built up from crystal structure information. QM/MM methods can straightforwardly be used for geometry optimisations of any system as long as the division of the QM and MM region is well-defined and a force field is available for the MM region. The QM region will contain the NMR nucleus and the geometry of the QM region will change as the QM atoms interact with the MM atoms, hence capturing some of the geometric environmental effects. Both mechanical and electrostatic embedding schemes can be used for such calculations although an electrostatic embedding method was always used in this work.
The second environmental effect, how the NMR property is directly influenced by the environment by perturbation of the electronic wavefunction, is only describable with electrostatic embedding and in fact the polarisation of the electron wavefunction/density by the surrounding point charges is the only part of the MM region that will affect it. Hence, for single-point QM/MM calculations, only the point charges of the MM region need to be defined.

3.2 Evaluation of DFT methods for ^{51}V chemical shift calculations

One of our goals is the accurate computation of anisotropic transition metal chemical shifts in the solid state with special emphasis on ^{51}V NMR (see Chapters 3.4, 3.5.4, 3.5.5 and 3.6). The solid state introduces environmental effects of unknown magnitude but it is also clear that a large error in these calculations stems from the electronic structure method used to calculate the shielding tensor. The challenge is to distinguish between errors resulting from the neglect or approximations of the environment and intrinsic errors resulting from the choice of the exchange-correlation functional and basis set.

If experimental ^{51}V chemical shifts of vanadium compounds in the gas-phase were available then they would be natural reference values to compare to, enabling one to see how well current electronic structure methods can predict such chemical shifts, free from intermolecular interactions. Gas-phase NMR spectroscopy remains a specialised field, however, and results for only a few compounds are available and to our knowledge none for transition metal compounds. Liquid state ^{51}V NMR is very common, however, and a considerable amount of data is available.[125] In a benchmark study,[126] a test set of vanadium complexes was compiled from various parts of the ^{51}V chemical shift range for which experimental solution chemical shifts were available and a few methods tested. The GGA functional BP86 and the hybrid B3LYP were both found to perform well, especially considering the large chemical shift range of ^{51}V compounds, 4600 ppm. A mean absolute error of 118 ppm with respect to experiment with the B3LYP functional (and the GIAO method) and a linear regression

slope slightly larger than unity was found. B3LYP has become the most used functional for ^{51}V NMR calculations as it seems generally reliable for transition metal chemical shifts.

Since this study, several new density functionals have become available in quantum chemistry codes, including several functionals of the meta-GGA and hybrid meta-GGA type. We set out to test several of these recently available meta-GGA and hybrid meta-GGA functionals for the computation of ^{51}V chemical shifts, using the same test set as the previous study to see if meta-GGA or hybrid-meta GGA functionals offer an improvement for the computation of ^{51}V chemical shifts over hybrid GGA functionals like B3LYP or BP86.

The test set, shown in figure 2, consists mostly of V(V) compounds with two molecules (**7** and **8**) with V in different oxidation states (negative oxidation states in fact). **9**-**16** are model compounds for systems with bulkier groups (OtBu groups instead of OMe groups) but the difference in the chemical shifts between the model compounds and the real compounds should be minimal as discussed in the original study.[126]

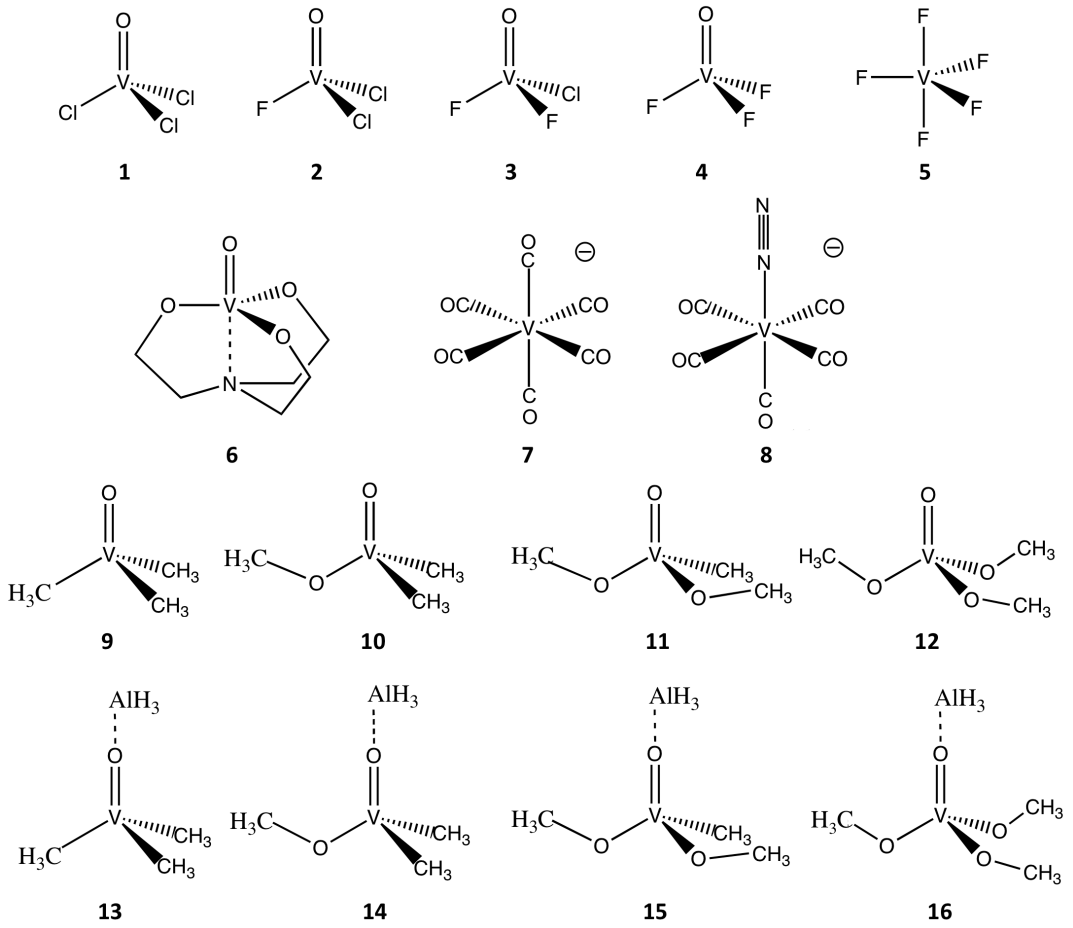


Figure 2 The test set of vanadium compounds with experimentally known ^{51}V chemical shifts.

The complex **1** is the reference compound in ^{51}V NMR spectroscopy and all chemical shifts of the compounds **2**-**16** were calculated as shielding differences according to Eq. 56, with σ_{ref} being the σ_{iso} of **1**, $VOCl_3$. The observed $\delta(^{51}V)$ values of this test set cover -1952 ppm to 1575 ppm, a large fraction of the ^{51}V chemical shift range.

Geometry optimisations and NMR calculations were performed using the quantum chemistry code Gaussian 09.[64] All single-point NMR calculations used the large def2-QZVPP basis set.[45] We tried decontracting this basis set but the results changed only marginally, suggesting that this basis set is close to the basis set limit for ^{51}V NMR calculations.

Four different setups were used to calculate chemical shifts:
Setup I: BP86/AE1 optimised geometries. Whole set.
Setup II: TPSS/QZVPP optimised geometries, Whole set.
Setup III: TPSS/QZVPP optimised geometries, subset (**1**-**9**, **13**).
Setup IV: TPSS/QZVPP optimised geometries, subset (**1**-**9**, **13**) with CPMD dynamical corrections.

For setup IV, thermal effects on the chemical shifts were explored by adding corrections from a Car-Parrinello molecular dynamics study[127] of chemical shifts of transition metal complexes. The difference between the averaged chemical shift from a BP86 CPMD trajectory and the chemical shift at the equilibrium geometry, ($\delta^{CPMD}{}_{ave}$ - δ_{eq}), were added as corrections for each of the complexes in setup IV (taken from Table 7 in Ref.[127]).

Table 1 contains the mean absolute error (MAE) as well as the slope of linear regression of calculated chemical shifts vs. experimental shifts, both criteria having been used in the previous study to judge the capability of functionals to predict chemical shifts.

Table 1 Mean absolute errors (ppm) and slopes of linear regression for different functionals for the test sets of ^{51}V chemical shifts. n.c. : not calculated.

Functional[a]	Setup I	Setup II	Setup III	Setup IV
B3LYP (20 %)	133.7 (1.04)	112.9 (1.05)	139.5 (1.07)	130.1 (1.06)
B3LYP (15 %)	n.c.	96.3 (1.03)	115.9 (1.04)	106.5 (1.03)
B3LYP (10 %)	n.c.	96.1 (1.00)	112.5 (1.01)	96.8 (1.01)
B3LYP (5 %)	n.c.	96.7 (0.98)	110.2 (0.99)	94.5 (0.98)
BLYP	n.c.	104.5 (0.95)	115.2 (0.97)	92.7 (0.96)
LC-ωPBE[128,129] **(lr. 100 %)**	168.2 (1.01)	146.7 (1.02)	152.6 (0.99)	144.3 (0.98)
BB95[130]	105.9 (0.98)	93.7 (0.98)	100.8 (1.00)	77.5 (0.99)
BB95 (5 %)	n.c.	81.4 (1.01)	93.8 (1.03)	78.1 (1.02)
BB95 (10 %)	n.c.	75.5 (1.04)	94.9 (1.06)	84.6 (1.05)
BB95 (15 %)	n.c.	76.5 (1.07)	100.7 (1.09)	103.2 (1.07)
B1B95 (28 %)[b]	156.8 (1.14)	149.2 (1.14)	192.6 (1.16)	192.1 (1.14)
BMK[131] **(42 %)**	256.7 (1.20)	254.6 (1.22)	311.7 (1.24)	311.8 (1.23)
τ-HCTH[132]	126.7 (0.95)	115.5 (0.95)	124.8 (0.97)	101.6 (0.96)
τ-HCTHh (5 %)	n.c.	91.9 (0.98)	102.7 (1.00)	77.4 (0.99)
τ -HCTHh (10 %)	n.c.	81.6 (1.00)	93.1 (1.03)	77.4 (1.01)
τ -HCTHh (15 %)[c]	109.4 (1.03)	93.3 (1.04)	116.2 (1.06)	106.7 (1.05)
TPSS[133]	146.7 (0.89)	129.8 (0.91)	142.1 (0.93)	110.9 (0.92)
TPSSh[133] **(10 %)**	127.6 (0.93)	117.3 (0.96)	131.2 (0.98)	115.5 (0.97)
PBE[134]	n.c.	113.1 (0.95)	121.5 (0.9)	90.6 (0.96)
BP86[19]	n.c.	108.2 (0.96)	116.4 (0.97)	87.0 (0.96)
B3P86 (20 %)	n.c.	99.1 (1.06)	126.4 (1.08)	118.2 (1.07)

[a] Functionals with the amount of Hartree-Fock exchange in hybrids in parentheses. [b] Recommended BB95 hybrid.[130] [c]Recommended τ –HCTH hybrid.[132]

Over 25 different functionals were tested for Setup I. The results for only a few are shown in Table 1. Dramatic effects of including Hartree-Fock exchange can be seen in the results. The worst-performing functionals are the ones that have a high percentage of Hartree-Fock exchange, e.g. BMK and the range-separated hybrid LC-ωPBE. Hybrid functionals with a small percentage (10-15 %) of Hartree-Fock exchange perform much better but do not always result in an improvement over their parent GGA functionals.

Two functionals, BB95 and the τ-HCTH-hybrid (15 %) were found to result in lower mean absolute errors than all other functionals. These two functional forms were thus further explored (as well as B3LYP) by varying the amount of Hartree-Fock exchange in their hybrid form to see the effect on the chemical shifts.

Results using setup II use geometries calculated at the TPSS/def2-QZVPP level. In a benchmark study of first-row transitional metal complex equilibrium geometries, the TPSS functional was found to be an improvement over BP86 for experimental metal-ligand bonds and the best performing functional in the study.[135] A large def2-QZVPP basis set should furthermore result in a minimal basis set error on the geometries. NMR results using TPSS/def2-QZVPP geometries show some improvement over BP86/AE1 geometries for all tested functionals, revealing that a part of the error in ^{51}V chemical shift calculations comes from geometrical effects.

Setup III uses TPSS/def2-QZVPP geometries and is a subset of the test set in Figure 2 showing slightly different chemical shift statistics than Setup II. When including the available CPMD dynamical corrections, i.e. setup IV, we can see that there is an improvement when using the dynamical corrections, especially when combined with

the BB95 functional, low percentage BB95 hybrids and low percentage τ-HCTH hybrids. A comparison of Figure 3 and 4 reveals this and shows in particular how complexes **7**, **8** and **13** benefit from the dynamical corrections which improve the overall agreement with experiment.

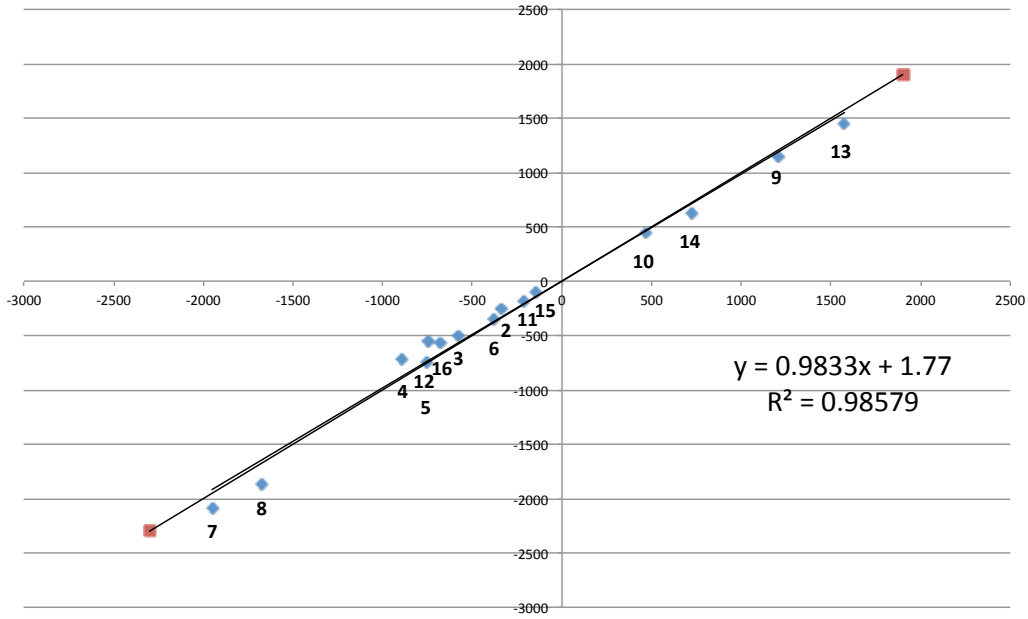


Figure 3 Computed vs. experimental chemical shifts (Setup II) with the BB95 functional. The line between the squares is the ideal line and the other is the linear regression line.

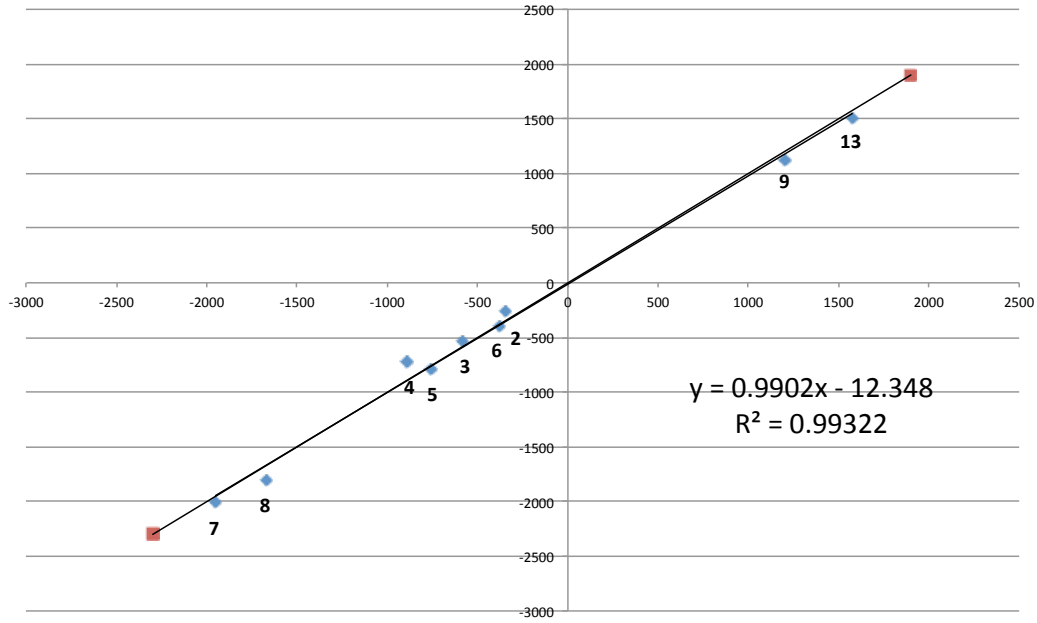


Figure 4 Computed vs. experimental chemical shifts (Setup IV) with the BB95 functional. The line between the squares is the ideal line and the other is the linear regression line.

It thus seems that the meta-GGA functionals BB95 and the τ-HCTH with a small or no amount of Hartree-Fock exchange are an improvement over the B3LYP functional. It is also evident that accurate geometries and dynamical corrections (if affordable) will usually improve agreement with experiment. Range-separated hybrids and

functionals with a large percentage of Hartree-Fock exchange result in much worse ^{51}V chemical shifts. This is consistent with recent studies reporting no improvement of range-separated hybrid functionals over conventional functionals for transition metal chemistry test sets.[136]
For now, however, it must be acknowledged that the improvement over B3LYP for BB95/ τ-HCTHh functionals is small. A significant amount of B3LYP ^{51}V NMR data exists already and thus the recommendation of a new functional for ^{51}V chemical shift computations may be premature. We suggest future studies to compare functionals to larger test sets to confirm whether the BB95 and the τ-HCTH hybrid functionals are a reliable improvement over B3LYP, both for ^{51}V and potentially other transition metal chemical shifts possibly taking into account solvent effects as well as they could have a small effect. We also found that a lower % hybrid form (5-10%) of B3LYP gave slightly better statistics than the normal 20 % form which should be explored further as well.

Furthermore, we note that isotropic chemical shifts are the average of the principal components of the symmetric NMR tensor. There have been studies that suggest that comparing only isotropic chemical shifts in benchmarking studies may conceal the true errors of many functionals for describing the actual chemical shift tensor.[137] Additional insight into the performance of functionals may thus come from the comparison of computed and experimental anisotropic NMR data. Such data is usually only available from solid-state NMR experiments, however, where solid-state effects must ideally be directly taken into account. We also note that our results depend on the reference shielding value, i.e. $VOCl_3$. Since experimentally, liquid $VOCl_3$ is the NMR standard while computationally, gaseous $VOCl_3$ is the standard, a substantial gas-to-liquid shift would lead to a systematic error of all values, affecting our comparison. The gas-to-liquid shift of $VOCl_3$ is discussed in Chapter 3.4.4.

3.3 Gas-phase computations of electric field gradients

The quadrupole interaction is observable through a variety of experimental techniques. The most direct way is through nuclear quadrupole resonance spectroscopy (NQR), but it can also surface in microwave (MW), Mössbauer, electron-paramagnetic resonance (EPR), electron-nuclear double resonance (ENDOR) and, probably most importantly, NMR spectroscopy. The quadrupole interaction affects the lineshape of NMR signals by providing a relaxation mechanism for the nuclear spin states, resulting in quadrupolar line broadening.[91] This effect can be described by two independent parameters, the nuclear quadrupole coupling constant (NQCC) and an asymmetry parameter, which can both be refined from high-resolution anisotropic NMR spectroscopy, usually in the solid state.

An important part of first principles modelling of solid-state NMR properties of transition metal compounds is the calculation of quadrupole interaction parameters. Of the 3d transition metals, all the NMR-active nuclei are quadrupolar except for ^{57}Fe (which is, however, Mössbauer active which results in quadrupolar splitting of Mössbauer spectra).

The quadrupole interaction is a sensitive probe into the electronic structure as it depends on the electronic and nuclear charge distribution around the quadrupole, through the electric field gradient (EFG). The EFG can thus be used to probe structural and chemical features of a molecule or solid whenever the quadrupole interaction parameters can be obtained.

One of the goals of this PhD work is the accurate calculation of the EFG tensor in the solid state. This will require an accurate approach to accounting for environmental effects on the geometry and the property itself. However, a crucial aspect will always be the electronic structure problem and the computation of EFG tensors has not been assessed as rigorously as chemical shift tensors.

A comparison between the results from the electronic structure calculations and experiment will be necessary but most experimental studies of the quadrupole interaction come from the solid state where the environment may influence the quadrupole interactions substantially. As we seek to explore the accuracy of the electronic structure calculation itself, a comparison with gas-phase experimental results would thus be ideal, i.e. “uncontaminated” by environmental effects. This would then enable one to find the best performing DFT functional or wavefunction method without worrying about the environment. Even though the EFG is a simple first-order property and is just calculated directly as an expectation value, it has been found that highly correlated wavefunctions and large basis sets may sometimes be needed for accurate results.[61,138,139] For transition metal complexes, DFT is often the only practicable quantum chemical method, MP2 generally performs poorly[98-100] and higher correlated methods, like coupled cluster methods, usually are not affordable. Quantum chemical calculations of EFG tensors of transition metal complexes and solids are starting to become common,[101,140] but only a few studies have systematically explored the accuracy of the methods that are in use, and these studies have mostly focused on small diatomics. Schwerdtfeger et al. showed that for simple diatomic first-row transition metal (especially copper) compounds, EFG tensors calculated using several common DFT functionals are greatly in error (even yielding the wrong sign of V_{zz}), when compared to both experimental and CCSD(T) data.[141-143]

In contrast, calculated EFG tensors of main-group compounds compare generally well with experiment.[61]

As previous studies had mostly focused on diatomic metal compounds (and on the troublesome group 11 of the periodic table, where relativistic effects are particularly pronounced), we felt a study on the performance of DFT methods for calculating EFG tensors of larger metal complexes with saturated ligand shell and less exotic electronic structures was needed, in particular for transition metals where the quadrupole coupling shows up in solid-state NMR spectroscopy. Such a study allows one to make a better estimate of the size of errors to anticipate from EFG calculations of transition metal complexes with DFT, help in deciding what functional to choose and hopefully shed some light on why transition metal EFG tensors seem to be harder to calculate accurately than main group EFG tensors.[61]

3.3.1 Benchmarking study of EFG tensors: assessment of functionals, basis sets and geometries

As we wanted to exclude environmental effects from our comparison we created a test set of first-row transition-metal complexes, for which accurate transition metal NQCCs are available from gas-phase microwave (MW) spectroscopy. The test set is shown in Figure 5. All NQCC data were converted into "experimental" V_{zz} EFG values using Eq. 49 and the latest recommended quadrupole moments from Pekka Pyykkö.[144] The following study was published in *Dalton Transactions* in 2010.[145]

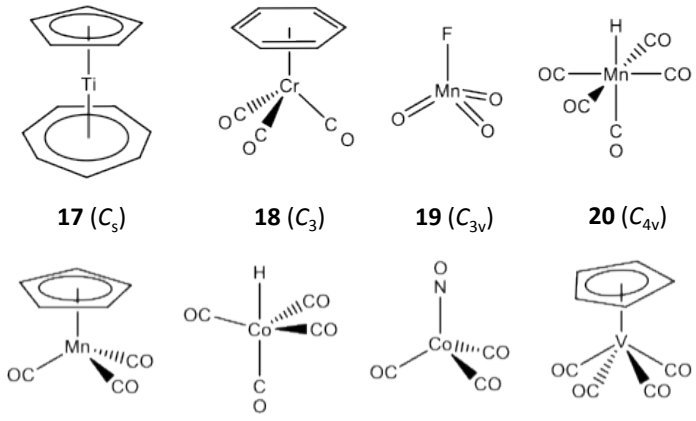


Figure 5 The test set of metal complexes with quadrupole coupling data available.

BP86/AE1 optimised geometries were used for the most part but the sensitivity towards the level of geometry optimisation was also separately tested. A "fine" integration grid (Gaussian keyword) was mainly used for numerical integration of the exchange-correlation term and an "ultrafine" grid in some cases (results were found to be insensitive to the grid for all contracted basis calculations). Using the large def2-QZVPP basis set, electric field gradients were calculated for the metal complexes using several different DFT functionals and compared to the experimentally derived EFGs. Most calculations were performed with Gaussian 09[64] while some calculations were performed with NWChem version 6.0[146] and Orca version 2.9[147] depending on availability of functionals in programs. The numerical accuracy (mainly controlled by

the SCF convergence criterion and the integration grid) between programs was checked. The results are shown in Table 2.

The initial comparison of computed vs. experimental data revealed an unusually large discrepancy for the Ti complex **17**. The computed V_{zz} values were larger than the experimental one by roughly a whole order of magnitude. Correspondence with the experimental scientists behind the original study[148] prompted a reinvestigation of the original measurements. The reinvestigation later revealed the original NQCC to be incompatible with the observed quadrupole splittings in the new MW spectrum. However, a new NQCC could not be refined due to uncertainties in line assignments. A NQCC of +57 MHz (from DFT calculations; V_{zz}= -0.8 au)[149] was found to fit the new data much better which would be in good agreement with our DFT results. The data for **17** was removed from all statistical analysis when comparing to experiment and will not be discussed further.

Table 2 EFG calculations with different functionals: V_{zz} values in au,[a] compared to gas-phase MW data.

Comp.	LDA	BP86	PBE	TPSS	B3LYP	PBE0	TPSSh	M06[24]	LC-ωPBE	RI-B2-PLYP[150]	Exp.
17	-0.790	-0.726	-0.714	-0.719	-0.775	-0.765	-0.737	-0.558	-0.866	-0.781	-0.115[148]
18	-0.348	-0.326	-0.316	-0.322	-0.339	-0.329	-0.328	-0.388	-0.433	-0.600	-0.344[151]
19	-0.569	-0.567	-0.554	-0.514	-0.451	-0.395	-0.465	-0.511	-0.280	-0.115	-0.217[152]
20	0.500	0.502	0.495	0.499	0.533	0.544	0.514	0.565	0.566	0.555	0.570[153]
21	-0.782	-0.758	-0.730	-0.742	-0.812	-0.817	-0.773	-0.874	-0.910	-0.839	-0.877[154]
22	-0.811	-0.880	-0.869	-0.926	-1.123	-1.165	-1.082	-1.258	-1.164	-0.407	-1.182[155]
23	-0.423	-0.401	-0.400	-0.398	-0.478	-0.520	-0.442	-0.538	-0.581	-0.466	-0.356[156]
24	-0.378	-0.340	-0.337	-0.331	-0.294	-0.281	-0.326	-0.331	-0.289	-0.211	-0.390[157]
MAE[b]	0.139	0.136	0.142	0.126	0.088	0.081	0.096	0.095	0.076	0.211	
ME[b]	-0.002	0.004	0.012	0.009	-0.024	-0.024	-0.015	-0.077	-0.042	0.102	
MaxE[b]	-0.371	0.351	0.337	0.297	0.234	0.178	0.248	0.295	0.225	-0.775	
Slope[b]	0.741	0.767	0.753	0.787	0.912	0.942	0.874	0.994	0.985	0.649	

[a]Gaussian-style values, compatible with eq. 49. Def2-QZVPP basis set and BP86/AE1 geometries.
[b]Mean absolute error (MAE), mean error (ME), maximum error (MaxE) and slope from a linear regression of DFT vs. experimental data (excluding complex **17**, see text).

The EFG results with the local density approximation (LDA) are interestingly in reasonable agreement with experiment, much better than Hartree-Fock data (MAE=0.32 au[145]). Curiously, the generalized-gradient approximation (GGA) does not seem to improve upon LDA. The BP86 functional is marginally better while the PBE results are even slightly worse (compare, for instance, the mean absolute errors, MAE, in Table 2). The meta-GGA approximation exemplified by the TPSS functional performs slightly better than LDA. The (global) hybrid functionals B3LYP, PBE0 and TPSSh, on the other hand, give considerably better results than their parent GGA variants. We also considered the newer range-separated hybrid functionals. The long-range corrected LC-ωPBE, that has the correct asymptotic behaviour of the exchange potential is the best performer of all the DFT methods. The parameterised range-separated hybrid CAM-B3LYP[158] gives similar results (see Ref.[145]).

The chemical shift tensor is known to be sensitive to geometrical effects. In order to probe for the geometry dependence of EFGs, we evaluated them at the

B3LYP/def2-QZVPP level using geometries from different sources (Table 3). Changing the geometry from BP86/AE1 to TPSS/AE1 has negligible effects on the MAE, while slightly improving the slope. Further enlarging the basis set used in the TPSS optimisations to def2-TZVPP and def2-QZVPP on all atoms has also very small effects on the EFG. Using the experimental geometries from either MW or GED studies, rather than DFT optimised ones, results in small changes in most cases.

Table 3 B3LYP/QZVPP V_{zz} values (in au), computed for different computed and experimental MW geometries.[a]

Geo. opt. level	BP86/ AE1	TPSS/ AE1	TPSS/ TZVPP	TPSS/ QZVPP	Expt.	Ref.
18	-0.339	-0.324	-0.320	-0.322	-0.288	159
19	-0.451	-0.451	-0.451	-0.461	-0.489	160
20	0.533	0.560	0.534	0.538	0.117	153
21	-0.812	-0.802	-0.804	-0.805	-0.793	161
22	-1.123	-1.129	-1.154[c]	-1.154	-1.253	155
23	-0.478	-0.465	-0.466[d]	-0.452	-0.399	162
24	-0.294	-0.273	-0.257	-0.256	-0.252	163
MAE[b]	0.088	0.088	0.091	0.090	0.159	
ME[b]	-0.024	-0.013	-0.017	-0.016	-0.080	
MaxE[b]	0.234	0.235	0.234	0.244	0.453	
Slope[b]	0.912	0.927	0.925	0.926	0.741	

[a] Using experimental gas-phase geometries from the given references. [b]See footnote b in Table 2.

For the Mn hydride complex **20**, however, the result is significantly worse, producing an error of 0.45 au, which causes the overall statistics to deteriorate notably (see ME and MAE values). In the case of **20** the optimised geometries, e.g. *r*(Mn-H), *r*(Mn-C^{ax}) and *r*(Mn-C^{eq}) of 1.58 Å, 1.85 Å and 1.85 Å, respectively (TPSS/QZVPP level), are probably more accurate than the MW-derived geometry, which is also associated with larger uncertainties, cf. *r*(Mn-H), *r*(Mn-C^{ax}) and *r*(Mn-C^{eq}) of 1.65(2) Å, 1.82(2) Å and 1.86 Å, respectively. Overall, however, EFG tensors at metal centres seem to be quite insensitive to geometrical details.

The def2-QZVPP basis set on all atoms is quite large and was at the time of this study assumed to be close to the DFT basis set limit (see Chapter 3.3.2 for a challenge of this assumption). We searched for more economical basis-set combinations and carried out additional EFG calculations with the B3LYP functional. The results obtained with a variety of basis sets are summarised in Table 4. With additional diffuse functions on all atoms (from the aug-cc-pVQZ basis set[46,48]), B3LYP/aug-QZVPP V_{zz} values were used as reference. Judging from the MAE values, a quadruple-zeta basis set on the metal atom is important for results close to the B3LYP/aug-QZVPP reference values, but a small double-zeta basis set like 6-31G* is mostly sufficient for the ligand atoms. The mixed QZVPP/6-31G* basis set is not much larger than the AE1 basis (Wachters/6-31G*), containing just an additional 51 basis functions on the metal, but gives results much closer to the reference values. Despite being highly unbalanced, this mixed basis set thus emerges as a cost-effective compromise for EFG computations for metal centers in transition-metal complexes. A triple-zeta basis set on all atoms, like 6-311+G(2d,p), can have many more basis functions than the mixed QZVPP/6-31G* basis set, but gives unsatisfactory results. As a rough guide for the size of a given basis-set combination, the number of basis functions for a typical medium-

sized complex, **21** in this case, is included in Table 4 (see numbers in parentheses in the header). As expected, substituting the inner core electrons of the metal with an effective core potential gives unacceptable results for the EFG tensor for many of the metal complexes (results not shown).

Table 4 B3LYP-computed EFGs (V_{zz} values in au) using different basis sets,[a] together with the mean absolute error (MAE) with respect to aug-QZVPP (in parentheses: total number of basis functions for complex **5**).

Comp.	aug-QZVPP (1268)	QZVPP (877)	QZVPP/ TZVPP (511)	QZVPP/ SVP (279)	QZVPP/ 6-31G* (264)	Wachters/ 6-31G* (213)	TZVPP (475)	TZVPP/ 6-31G* (228)	6-311+ G(2d,p) (392)	cc-pVTZ (468)
17	-0.781	-0.775	-0.776	-0.798	-0.809	-1.056	-0.887	-0.904	-1.031	-0.774
18	-0.344	-0.339	-0.340	-0.330	-0.338	-0.402	-0.409	-0.409	-0.404	-0.304
19	-0.411	-0.451	-0.441	-0.338	-0.333	-0.392	-0.328	-0.342	-0.375	-0.335
20	0.523	0.533	0.522	0.522	0.529	0.621	0.550	0.554	0.649	0.577
21	-0.814	-0.812	-0.809	-0.802	-0.795	-0.960	-0.903	-0.909	-0.965	-0.809
22	-1.183	-1.123	-1.136	-1.134	-1.180	-1.384	-1.175	-1.170	-1.379	-1.163
23	-0.457	-0.478	-0.469	-0.477	-0.483	-0.649	-0.554	-0.548	-0.599	-0.521
24	-0.291	-0.294	-0.285	-0.269	-0.244	-0.323	-0.318	-0.325	-0.304	-0.233
MAE		0.018	0.014	0.027	0.026	0.128	0.063	0.065	0.122	0.040

[a]Notation "basis on metal/basis on ligands".

3.3.2 Re-assessment of EFG tensors: decontraction of basis sets

In the original study which was described in Chapter 3.3.1 and published in *Dalton Transactions*[145], the large def2-QZVPP basis set, with and without augmented diffuse functions, was assumed to be close to the basis set limit for the DFT functionals trialled. DFT calculations are known to be much less basis set sensitive than WFT methods and smaller triple-zeta basis sets are usually found to be close to the basis set limit for energies, geometries and many other properties. At the time, even larger correlation-consistent basis sets were trialled (see Electronic Supporting Information of the study[145]).

Later, we decided to revisit this issue. Basis sets such as def2-QZVPP and cc-pV*n*Z are contracted basis sets which means that each basis function consists of a linear combination of contracted Gaussian functions where the contraction coefficients are fixed at values that were found to result in accurate energies when the basis set was originally developed.
A molecular property such as an EFG may, however, be more sensitive and dependent on basis functions and contracted Gaussians, than relative energies which is how basis set convergence is often assessed. By decontracting the basis set, each contracted Gaussian function becomes a basis function on its own, hence resulting in much more flexibility of the total basis set. The SCF calculation now will have more parameters to variationally optimise which will according to the variational principle lead to both a lower energy solution and a more accurate wavefunction/electron density (within the limit of the approximate exchange-correlation term) with an increase in computational cost.

The size of the integration grid for numerical integrations of the exchange-correlation term had to be increased for these decontracted calculations. A grid of the Murray-

Handy-Laming type was necessary in order to avoid inaccuracies for core electron integrals (according to the Gaussian program error output). A grid specified by the Gaussian keyword "Int(Grid=-220020)", i.e. 2*220*20*20=176000 points per atom, was found to result in small integration errors (compared to results for a huge Int(Grid=-300100) grid, while not being overly expensive.

Decontraction was found to have a dramatic impact on the EFG parameters. Table 5 shows calculated EFG values using large decontracted and contracted basis sets. The largest decontracted basis set combination, aug-cc-pwCVQZ[164]/aug-cc-pCVQZ,[46,48,165] (that includes a number of extra core and core-valence basis functions), was now assumed to be the basis set closest to the basis set limit and was used as reference. The MAE with respect to the reference values (i.e. an estimate of the basis set error) is considerably different for normal def2-QZVPP (MAE=0.088) and decontracted def2-QZVPP (MAE=0.015) basis sets. This demonstrates that basis set errors of 0.09 au (which is even larger than previously estimated DFT functional errors) can be present in normal DFT/def2-QZVPP calculations and that normal contracted basis set calculations are unsuitable for reliable EFG calculations. It was also found that the decontraction is only crucial for the metal atom and the results suggest that a normal contracted double-zeta basis set on the ligand atoms can give very good results as long as a large decontracted basis set is used on the metal atom.

Table 5 PBE-computed EFGs (V_{zz} values in au) using different basis sets combinations together with the mean absolute error (MAE) with respect to aug-cc-pwCVQZ and experiment.

	Experiment	aug-cc-pwCVQZ[a]/ aug-cc-pCVQZ[a]	cc-pVQZ[a]/ cc-pVQZ	def2-QZVPP	def2-QZVPP[a]	def2-QZVPP/ 6-31G*	def2-QZVPP[a]/ 6-31G*
18	-0.344	-0.362	-0.364	-0.316	-0.359	-0.327	-0.369
19	-0.217	-0.498	-0.499	-0.554	-0.479	-0.470	-0.486
20	0.57	0.591	0.592	0.495	0.572	0.493	0.574
21	-0.877	-0.862	-0.870	-0.730	-0.851	-0.736	-0.863
22	-1.182	-1.033	-1.048	-0.869	-1.056	-0.921	-1.081
23	-0.356	-0.512	-0.521	-0.400	-0.491	-0.398	-0.497
24	-0.39	-0.347	-0.346	-0.337	-0.337	-0.296	-0.341
	MAE_{basis}[b]	-	0.006	0.088	0.015	0.081	0.015
	MAE_{exp}	0.098	0.096	0.142	0.088	0.126	0.086

[a] Decontracted [b] Mean absolute error w.r.t. to the decontracted aug-cc-pwCVQZ/aug-cc-pCVQZ basis set.

The basis set error of decontracted cc-pVQZ on metal and normal cc-pVQZ on ligands is very small and we decided to use this basis set for a complete re-evaluation of DFT functionals for EFGs. A new functional comparison is shown in Table 6.

Table 6 Decontracted basis EFG calculations with different functionals: V_{zz} values in au,[a] compared to gas-phase MW data.

Comp.	LDA	BP86	PBE	TPSS	B3LYP	PBE0	TPSSh	revTPSS	revTPSSh	LC-ωPBE	Exp.
17	-0.982	-0.964	-0.944	-0.960	-1.010	-0.979	-0.973	-	-	-1.086	-0.115
18	-0.384	-0.372	-0.365	-0.360	-0.386	-0.381	-0.366	-0.350	-0.349	-0.494	-0.344
19	-0.542	-0.499	-0.499	-0.454	-0.348	-0.302	-0.383	-0.417	-0.347	-0.182	-0.217
20	0.607	0.598	0.592	0.594	0.634	0.638	0.611	0.566	0.583	0.666	0.570
21	-0.900	-0.881	-0.870	-0.861	-0.948	-0.953	-0.894	-0.834	-0.867	-1.059	-0.877
22	-0.985	-1.062	-1.048	-1.110	-1.336	-1.365	-1.226	-1.113	-1.229	-1.400	-1.182
23	-0.536	-0.527	-0.521	-0.504	-0.611	-0.642	-0.553	-0.483	-0.530	-0.714	-0.356
24	-0.385	-0.351	-0.346	-0.334	-0.301	-0.289	-0.313	-0.326	-0.299	-0.292	-0.390
MAE[b]	0.115	0.096	0.096	0.081	0.115	0.119	0.081	0.073	0.067	0.162	
ME[b]	-0.047	-0.045	-0.037	-0.033	-0.071	-0.071	-0.046	-0.023	-0.034	-0.097	
MaxE[b]	0.325	0.282	0.282	0.237	0.255	0.286	0.197	0.200	0.174	0.358	
Slope[b]	0.903	0.940	0.919	0.950	1.099	1.119	1.024	0.934	1.009	1.183	

[a]Gaussian-style, BP86/AE1 geometries. Basis set combination was: decontracted cc-pVQZ (metal) and normal cc-pVQZ (ligands).
[b]Mean absolute error (MAE), mean error (ME), maximum error (MaxE) and slope from a linear regression of DFT vs. experimental data (excluding complex **17**, see text).

This new functional comparison is in rather sharp contrast to the previous one as hybrid functionals and especially LC-ωPBE, are now strongly disfavoured when using a more flexible decontracted basis set (CAM-B3LYP also gave bad results, MAE=0.146 au). This can be seen in Figure 6 and 7. A correlation with Jacob's ladder of DFT methods is more obvious in Figure 7 (decontracted calculations) in going from LDA to GGA to meta-GGA functionals but hybrid functionals break the trend. Including 20-25 % Hartree-Fock exchange results in large deviations while 10 % like in TPSSh seems OK. A recently developed, revised TPSS functional, revTPSS,[166] available in the Orca code was also tried out. A hybrid version (not previously used in the literature) with 10 % Hartree-Fock exchange was also tested (named revTPSSh). These two functionals were found to result in the lowest MAE values w.r.t. experiment.

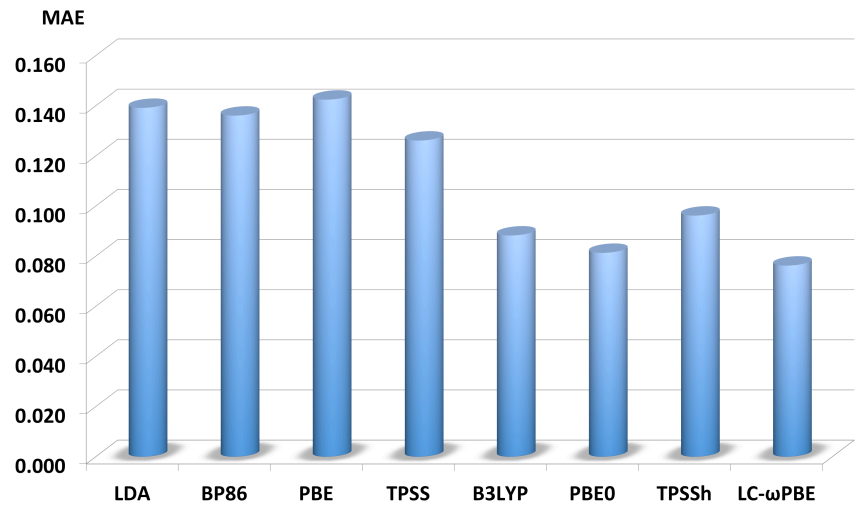


Figure 6 Plot of the mean absolute errors (in au) of different functionals for the EFG test set using the contracted def2-QZVPP basis set.

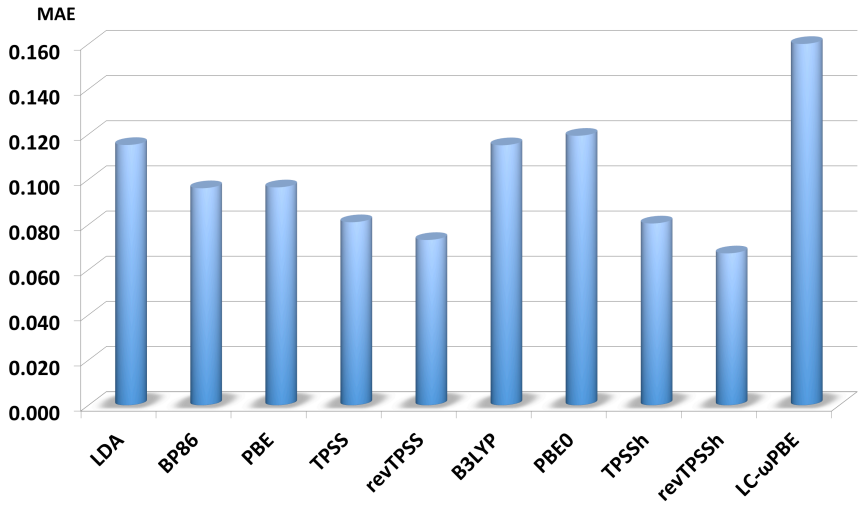


Figure 7 Plot of the mean absolute errors (in au) of different functionals for the EFG test set using the basis set combination cc-pVQZdecon/cc-pVQZ.

The revTPSS functional is a version of the TPSS meta-GGA functional where additional constraints of the density are satisfied (that the exact functional must have). This turned out to result in improved lattice energies and metal surface energies for revTPSS compared to TPSS while preserving the good behaviour for atomisation energies, compared to normal GGA functionals like PBE.[166]

Overall the TPSS, revTPSS, TPSSh and revTPSSh functionals result in the smallest devations with respect to the experimental values when decontracted basis sets are used and thus might be the best functionals to employ generally in EFG computations of transition metal complexes.

3.3.3 Relativistic effects

The Schrödinger equation is not a completely correct description of the interaction of electrons and nuclei as it ignores effects due to special relativity. The Dirac equation is a relativistically correct equation on the other hand and the Schrödinger equation can be understood as an approximation to it by taking the speed of light to be infinite.[2] Due to the complexity of the Dirac equation it is rarely affordable but for the smallest molecules, but popular approximations to it have been derived that can capture most of the effects of relativity.

Relativistic effects are known to become important to account for, for the lower transition metal rows, but are often ignored for the 1st transition metal row (copper being an exception). Relativistic effects have previously been shown to be negligible for ^{57}Fe EFG tensors[167] according to Zeroth-Order Regular Approximation (ZORA) [168,169] calculations, an approximation the relativistic Dirac equation. [170,170170] Recent developments in Douglas-Kroll-Hess relativistic approximations[171] that correct for the picture-change effect on electric field gradients[170] (an artifact arising due to the mismatch between a nonrelativistic operator being used with a quasirelativistic wavefunction/density), have allowed us to explore this issue by performing Douglas-Kroll-Hess (DKH) relativistic calculations using the Orca code.[147]

In a recent study, DKH electric field gradients were calculated for the hydrogen halides.[170] By using spin-free Dirac-DFT results as reference values for the molecule

test set, the convergence and performance of the DKH approximation could be systematically explored. The dependence of the results on the *n*-th order of the DKH*n* Hamiltonian as well as the order of the picture-change correction were explored and it was found that DKH2 (second order DKH) with picture-change corrections was a generally accurate approximation to capture scalar relativistic effects on EFGs. The picture-change correction was found to be crucial for accurate EFG calculations.

We carried out such DKH*n* calculations (and non-relativistic calculations for comparison) using the revTPSSh functional and the decontracted def2-QZVPP basis set using the Orca code. DKH*n* calculations up to *n*=3 were carried out. Calculations were carried out with and without picture-change corrections. The DKH2 calculations were found to be ~1.5 times more expensive than nonrelativistic calculations while DKH3 calculations (including picture-change corrections) were ~1.7 times more expensive.

Table 7 Decontracted basis EFG calculations with the revTPSSh functional and different relativistic approximations: V_{zz} values in au,[a] compared to gas-phase MW data.

Comp.	NR[b]	DKH1	DKH1-pc[c]	DKH2	DKH2-pc[c]	DKH3-pc[c]	Exp.
18	-0.355	-0.358	-0.362	-0.358	-0.362	-0.362	-0.344
19	-0.353	-0.372	-0.378	-0.372	-0.378	-0.378	-0.217
20	0.586	0.589	0.588	0.589	0.588	0.588	0.570
21	-0.871	-0.880	-0.880	-0.880	-0.880	-0.880	-0.877
22	-1.226	-1.201	-1.203	-1.203	-1.204	-1.204	-1.182
23	-0.532	-0.520	-0.534	-0.520	-0.535	-0.535	-0.356
24	-0.303	-0.317	-0.320	-0.317	-0.319	-0.319	-0.390
MAE[d]	0.068	0.064	0.067	0.064	0.067	0.067	
ME[d]	-0.037	-0.038	-0.042	-0.038	-0.042	-0.042	
MaxE[d]	0.176	0.164	0.178	0.164	0.179	0.179	
Slope[d]	1.008	0.999	0.998	1.000	0.999	0.999	

[a]Gaussian-style, decontracted def2-QZVPP basis set and BP86/AE1 geometries.
[b] NR: Non-relativistic.
[c] Picture-change correction used.
[d]Mean absolute error (MAE), mean error (ME), maximum error (MaxE) and slope from a linear regression of DFT vs. experimental data.

Table 7 shows the results of these calculations. The results reveal that there is sometimes an overestimation of relativistic effects when not including the picture change correction. Overall, the picture-change corrected relativistic effect is rather small and does not significantly affect the mean absolute error of EFGs for our test set (although some molecules seem to be more sensitive than others). This suggests that relativistic calculations for 3d transition metal EFGs are generally not needed as the DFT functional error will overshadow any error resulting from the neglect of relativistic effects.

3.3.4. Summary

We can summarise the results for gas-phase EFG calculations as follows:

1. The electric field gradient in the gas-phase does not seem overly sensitive to geometric effects (BP86/TPSS geometry optimisations with a moderate basis set (AE1) seem to be sufficient).
2. There is a strong basis set dependence and normal contracted basis sets are unreliable for EFG computations. In order to minimise basis set errors in EFG calculations, a large decontracted quadruple-zeta basis set on the metal atom is recommended, while much smaller (double-zeta) normal contracted basis sets can be used for the ligand atoms with good accuracy.
3. Relativistic effects on EFGs can be easily checked for nowadays as the additional computational cost is not excessive. Relativistic effects seem to be small for 1st transition metal row EFGs and can probably be ignored for the most part.
4. The DFT functional is the greatest source of uncertainty in these calculations and most likely responsible for the remaining deviations with experiment. The (rev)TPSS and (rev)TPSSh hybrid functionals seem to be the most reliable functionals for EFG computations of 3d transion metal complexes.
5. Finally it must be noted that this test set of EFGs is rather small and any error in the experimental coupling constant or quadrupole moment can significantly skew the results. The experimental quadrupole coupling constants have small error bars, however, some of the quadrupole moments have sizeable uncertainties, e.g. over 30 % for ^{53}Cr.
   Recently a new quadrupole moment of -4.8 fm^2 instead of -5.2 fm^2 has been suggested for ^{51}V.[172] We note, however, that using this value for Q would increase the deviation for the EFG at the V nucleus for all tested functionals as all of them underestimate the absolute value for this EFG while the new Q would increase the absolute value. Hence, this would affect the comparison between methods very little.
6. Our DFT functional recommendations should be further explored in the future. It is likely that test sets from solid-state experiments are necessary as microwave NQCC determinations for transtion metal complexes are few. This will inevitably require reliable methods to take solid-state effects into account.

Finally, we give a table with estimated DFT functional errors for selected first transition metal row nuclides (in MHz) based on the computed mean absolute errors in EFG V_{zz} values in Table 6 and Eq. 49. Later we will compare computed and experimental ^{51}V and ^{59}Co NQCC data in the solid-state where this table will become useful.

Table 8 Estimated DFT errors for NQCC (in MHz) of the first transition metal row nuclides for selected functionals.[a]

Comp.	B3LYP	PBE	TPSS	TPSSh	revTPSS	revTPSSh
^{45}Sc	6.0	5.0	4.2	4.2	3.8	3.5
^{47}Ti	8.2	6.8	5.8	5.7	5.2	4.8
^{49}Ti	6.7	5.6	4.7	4.7	4.2	3.9
^{51}V	1.4	1.2	1.0	1.0	0.9	0.8
^{53}Cr	4.1	3.4	2.9	2.8	2.6	2.4
^{55}Mn	8.9	7.5	6.3	6.2	5.7	5.2
^{57}Fe	4.3	3.6	3.0	3.0	2.8	2.5
^{59}Co	11.4	9.5	8.0	7.9	7.2	6.6
^{61}Ni	4.4	3.7	3.1	3.1	2.8	2.6
^{63}Cu	6.0	5.0	4.2	4.2	3.8	3.5
^{65}Cu	5.5	4.6	3.9	3.9	3.5	3.2
^{67}Zn	4.1	3.4	2.9	2.8	3.5	2.4

[a] Mean absolute NQCC errors for each nuclide, calculated according to Eq. 49, using the mean absolute error of V_{zz} from Table 6 as V_{zz} value and Q for each nuclide is taken from the latest set of quadrupole moments from Pyykkö.[144]

3.4 Computing ^{51}V NMR parameters of $VOCl_3$ in the solid-state

The majority of NMR computations in the literature are gas-phase calculations where the environment is assumed to be negligible. This may often be a good approximation when comparing to liquid state NMR data obtained in inert solvents, but it is less clear how reliable gas-phase calculations of molecules are as approximations to the real molecular solids.

$VOCl_3$ (**1**) is a simple closed shell vanadium(V) complex. Liquid $VOCl_3$ is the reference standard in ^{51}V NMR spectroscopy. ^{51}V solid-state NMR studies of solid $VOCl_3$ have resulted in available anisotropic chemical shift (and an isotropic chemical shift relative to liquid $VOCl_3$) and quadrupole coupling parameters.[173] Solid-state ^{51}V NMR studies of vanadium(V) compounds are becoming increasingly popular in the context of vanadium-based catalysts and materials.

Due to the availability of these data for a simple but important vanadium complex, we decided to explore generally how a solid environment affects the NMR properties of such a simple molecule and to come up with a protocol to model environmental effects on NMR properties. In this work we made use of both periodic DFT calculations as well as QM/MM methods for the solid-state NMR calculations. This work was part of a larger study where both gas-to-liquid and liquid-to-solid chemical shifts of $VOCl_3$ were modelled by elaborate Car-Parrinello MD simulations, to be discussed in Chapter 3.4.4. The whole study was published in *Phys. Chem. Chem. Phys.* in 2011.[174]

Periodic DFT calculations enable one to perform geometry optimisations of the full crystalline solid and the use of QM/MM methods for the NMR calculations gives a greater variety of electronic structure methods (compared to periodic DFT methods) such as hybrid DFT methods. Additionally the use of single-point QM/MM methods for NMR calculations can completely avoid the use of pseudopotentials, a necessity in planewave periodic DFT approaches which can have detrimental effects on molecular properties that depend on core electrons.

The problem with using QM/MM methods for a transition metal complex such as $VOCl_3$ is that reliable force fields to describe the MM region are seldom available. While force field parameters could be developed on a per case basis this severely limits the ease with which such calculations can be performed. For this study, only point charges were defined for $VOCl_3$ which were used for electrostatically embedded single-point calculations only but no other force field parameters were defined which would be necessary for full QM/MM geometry optimisations.

3.4.1 Computational details

Geometry optimisations of solid $VOCl_3$ was performed using the planewave periodic DFT code CPMD[175] at the the BP86-D level, using the empirical dispersion correction by Grimme.[26] Modelling solid-state systems requires initial structural data, which for metal complexes usually comes in the form of X-ray crystal structures. Two sets of crystal structures for $VOCl_3$ were considered, by Galy et al. from 1983[176], and by Troyanov from 2005[177]. The latter has a slightly larger volume of the unit cell (by 2.5%) and a slightly lower R-value (reliability factor, a measure of how the model reproduces the diffraction data)[178] than the former (3.2% vs. 4.9%). Single-point energy computations found the Troyanov structure to be lower in energy and the Grimme dispersion correction was found to be crucial for a qualitatively correct cohesion energy.[174] A supercell (crystallographic cell from Troyanov[177] doubled along the a-axes) with dimensions of 10x9x11 $Å^3$ containing 8 molecules was calculated. As optimisations of unit cell parameters are difficult, they were not attempted. Norm-conserving Troullier-Martins pseudopotentials[179] were used in the Kleinman-Bylander form[180] and Kohn-Sham MOs were expanded at the Γ-point in a basis of plane waves with a cutoff of 80 Ry. These calculations were performed by Michael Bühl and more details are available in the published study.[174]

The accuracy of solid-state X-ray geometries vs. solid-state DFT-optimised geometries in NMR calculations has been discussed in several studies.[181-183] It seems that due to resolution limitations of the X-ray structures, geometry optimisations are generally preferable, especially for hydrogen positions. The geometry from a periodic BP86-D calculation on the Troyanov structure was eventually chosen as the most reliable crystal structure of $VOCl_3$.

The Chemshell program v. 3.3[88,89] was used to create a large spherical MM cluster with radius ~48 Å (90.0 au), from the DFT-optimised crystal structure, centered on the V atom of one $VOCl_3$ molecule with all partial molecules at the surface deleted. A simple Coulomb point charge force field for each MM atom type was defined by using atomic NPA charges from a single molecule calculated at the B3LYP/QZVPP level. Additional charges surrounding the cluster were added to simulate the electrostatic potential of the infinite periodic system, using a (now outdated) procedure in Chemshell named make_3d_corrected_covalent_cluster (superceded by the construct_cluster procedure, see later).

A simple scheme to iterate the MM point charges self-consistently was devised (similar as loop A in Chapter 3.5.1):
A quantum mechanical (QM) cluster (single $VOCl_3$ unit or larger, see below) was selected from the middle of the classical cluster, its charges deleted, thus resulting in coordinates for both the QM cluster and coordinates and charges for the MM atoms.

The coordinates and point charges of the entire cluster were subsequently used for single-point embedded QM calculations of the QM cluster, using the Gaussian 09 program. From these single-point embedded calculations, new NPA atomic charges were obtained that were used to update the Coulomb force field. This results in a cycle that was iterated until the charges were self-consistent. Single-point NMR calculations were performed with different density functionals using several different basis set combinations (basis set on metal/basis set on ligand): Wachters/6-31G* (AE1), QZVPP / 6-31G* and QZVPP/QZVPP as well as a combination of decontracted def2-QZVPP on metal and normal 6-31G* on ligand atoms.

Due to the small size of $VOCl_3$, it was possible to increase the size of the QM cluster for the single-point QM/MM calculations. This then allows one to probe how the chemical shift and EFG tensors are affected by the short-range quantum mechanical polarisation (i.e. beyond the classical electrostatic effect by point charges). Three different cluster models were defined: a single molecule (I), this molecule with the 14 nearest neighbours (II) and finally a large cluster consisting of 65 molecules in total (III). The different cluster models are illustrated in Figure 8.

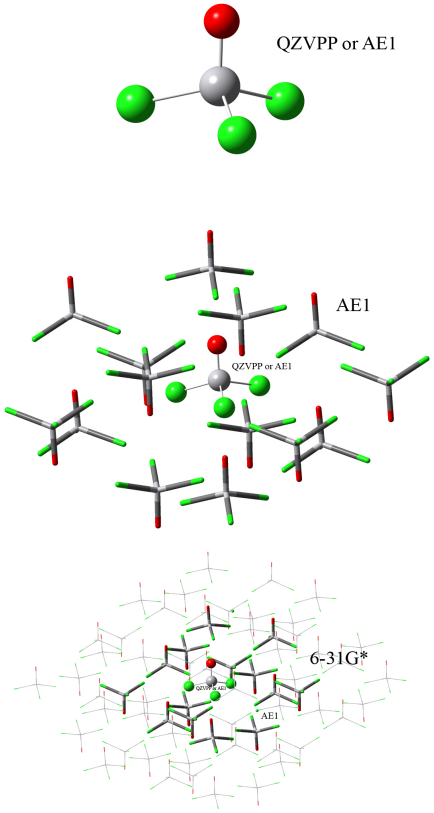


Figure 8 The three cluster models (from the top, clusters I, II and III) showing the different layers. Surrounding point charges not shown.

The single-point NMR calculations were performed with and without the surrounding point charges and subsequent charge updates. The charges converged quickly, only 2-3 iterations were necessary for the charges to stop changing more than 0.0001 e^-.

Different basis set were used for the central molecule: (notation: basis set on V/basis set on O and Cl) Wachters/6-31G* (AE1) and def2-QZVPP/def2-QZVPP. Due to the size of cluster models II and III, smaller basis sets were used for the added molecules

in the two additional layers. For the second layer (14 molecules) in cluster model II and III, the surrounding molecules were all assigned the AE1 basis set. In cluster model III, the additional molecules in the third layer used the 6-31G* basis set (for V, O and Cl). As the outer layers only serve the purpose of polarising the density of the central molecule, on which the NMR parameters are calculated, the use of smaller basis sets on the outer layers seems justified. Decontraction of the basis set of the central molecule was also explored as well as different basis set combinations including effective core potentials (ECP) in the outer layers.

3.4.2 Results: Converging the chemical shift and EFG tensors in the solid

Table 9 shows the computed NMR properties of the ^{51}V nucleus of the central molecule using the different cluster models. These results allow one to directly assess the effects of a solid-state geometry vs. a gas-phase geometry, the effect of cluster size and the effect of surrounding point charges on the ^{51}V chemical shift and EFG tensors.

Table 9 Calculated chemical shift and quadrupole coupling data with a single molecule (gas-phase optimised) model, and three different cluster models of the solid state (with and without point charges, PC). Shielding constants, σ_{iso} (ppm), gas-to-solid shifts (ppm), reduced shielding anisotropies, $\delta_\circ$ (ppm), C_Q values (MHz) and asymmetry parameters, η (unitless).

	Single molecule	**Cluster I**		**Cluster II**		**Cluster III**	
B3LYP/AE1[c]	Gas-phase opt.	Non-embed.	PC-Embed.[a]	Non-embed.	PC-embed.[a]	Non-emb.	PC-embed.[a]
σ_{iso}	-2267.6	-2268.1	-2269.4	-2287.2	-2286.6	-2286.8	-2287.5
Gas-solid shift, $\Delta\delta$	-	0.5	1.8	19.5	19.0	19.2	19.8
$\delta_\circ$	398.1	380.3	442.4	441.7	447.6	446.3	451.7
$\eta_\circ$	0.001	0.015	0.017	0.012	0.015	0.012	0.014
C_Q	7.95	7.58	7.38	7.31	7.28	7.29	7.26
η_Q	0.002	0.061	0.051	0.069	0.070	0.071	0.070
B3LYP/QZVPP[c]							
σ_{iso}	-2334.2	-2335.1	-2338.0	-2349.9	-2349.3	-2349.6	-2350.1
Gas-to-solid shift, $\Delta\delta$	-	1.0	3.8	15.7	15.2	15.4	16.0
$\delta_\circ$	354.3	337.3	401.7	403.4	409.7	409.2	414.8
$\eta_\circ$	0.001	0.016	0.020	0.017	0.020	0.017	0.019
C_Q	7.21	6.86	6.69	6.86	6.84	6.84	6.82
η_Q	0.002	0.066	0.055	0.071	0.072	0.073	0.073

[a] With surrounding point charges that have been updated twice.
[b] The gas-to-solid shift is calculated as $\Delta\delta = \sigma_{gas}-\sigma_{cluster}$.
[c] The basis set level of the central molecule (layer 1); layers 2 and 3 are described with AE1 and 6-31G* basis, respectively.

The different geometries and the surrounding point charges seem to have little effect on the shielding constant and, hence, the isotropic chemical shift, while a 15-20 ppm deshielding takes place when increasing the cluster size from one molecule to 15 or 65. This suggests that some environmental effects that directly affect the wavefunction/electron density cannot be described by simple point charges, i.e. an electrostatic description of the environment. A recent study of the shielding constants of liquid water showed this behaviour as well.[184] It thus seems necessary to include

the nearest neighbours for well converged shielding constants. The shielding constant is also very sensitive to the basis set level, as expected.

The anisotropic shift, on the other hand, shows a larger response to the point charges. There is a small effect in the reduced anisotropy, δ_σ , when using a solid-state geometry rather than a gas-phase geometry and a considerable effect when including point charges for the single-molecule clusters. When using the larger cluster model II we see that this shift in the anisotropy seems to be well modelled by the point charges, and only a small additional shift occurs when including point charges for cluster model II. Small additional changes occur when using cluster model III. A relative environmental effect on the shielding tensor thus can be modelled quite well by point charges it seems. The asymmetry parameter changes only a little in these calculations.

The quadrupole coupling constant, C_Q, shows a slight shift with the geometry change and some sensitivity toward the inclusion of point charges for cluster model I. This shift, however, is overestimated when comparing the B3LYP/QZVPP data for cluster model I and II. Only small additional changes occur when point charges are included for cluster model II and a similar change is found upon increasing the cluster to 65 molecules (cluster model III). Overall the quadrupole coupling constant does not change appreciably when comparing the gas-phase model with the large embedded cluster.

Point charge updates, all performed using cluster model I, have overall only small effects on the chemical shift and EFG tensors and converge quickly as can be seen in Table 10, suggesting small polarisation effects in the crystal whose intermolecular interactions are, arguably, dominated by dispersion. The largest effect is for the reduced anisotropy for the single-molecule cluster model. We also performed calculations using point charges derived from the embedded cluster model II, but tests using these charges resulted only in negligible changes, thus suggesting that gas-phase derived single-molecule point charges seem to be a fairly good approximation for this system and that the exact values of the point charges are not too important.

Table 10 B3LYP solid-state NMR data of cluster model I (BP86-D geometry) with embedded point charges, showing the effect of point charge updates.

		B3LYP/AE1	B3LYP/QZVPP
Run 1. With point charges from a gas-phase B3LYP/QZVPP calculation	σ_{iso}	-2269.1	-2337.4
	$\delta_{\circ}$	434.21	393.3
	$\eta_{\circ}$	0.016	0.017
	C_Q	7.41	6.72
	η_Q	0.053	0.057
Run 2: With point charges from embedded point charges from run 1.	σ_{iso}	-2269.4	-2337.9
	$\delta_{\circ}$	441.5	400.8
	$\eta_{\circ}$	0.017	0.020
	C_Q	7.38	6.69
	η_Q	0.051	0.055
Run 3: With point charges from embedded point charges from run 2.	σ_{iso}	-2269.4	-2338.0
	$\delta_{\circ}$	442.4	401.7
	$\eta_{\circ}$	0.017	0.0198
	C_Q	7.38	6.69
	η_Q	0.051	0.055

The convergence of NMR properties for this system was later (after publication in *Phys. Chem. Chem. Phys.*[174]) explored in more detail with respect to basis functions on the surrounding molecules of cluster model II. The results can be seen in Table 11. A very expensive calculation with the large def2-QZVPP basis sets for both layers was used as reference and the basis set in layer 2 was varied.

Table 11 Additional solid-state B3LYP NMR data with embedded cluster model II but with different basis sets[b] for layer 2.

Basis sets:								
Layer 1:	QZVPP	QZVPP	QZVPP	QZVPP	QZVPP	QZVPP	QZVPP	QZVPP
Layer 2:	QZVPP	AE1	6-31G*	CEP-121G	CEP-31G +pol[a]	CEP-31G	STO-3G	CEP-4G
# Bf.[b]	5400	1998	1788	1466	1340	1242	1060	1018
σ_{iso}	-2348.8	-2349.3	-2349.4	-2350.1	-2348.8	-2350.0	-2345.2	-2349.8
$\delta_{\circ}$	405.8	409.7	410.8	419.4	414.7	418.8	390.7	431.4
$\eta_{\circ}$	0.019	0.020	0.021	0.027	0.023	0.028	0.067	0.040
C_Q	6.91	6.83	6.78	6.85	6.86	6.84	6.86	6.78
η_Q	0.073	0.072	0.073	0.073	0.073	0.073	0.081	0.073

[a] An f polarisation function with exponent 0.257 on V was added (from the def2-QZVPP basis set).
[b] Number of basis functions for whole cluster.

In particular we tested basis sets which include effective core potentials in the outer layer, like the CEP basis set/ECP combinations (compact effective potential,[185]): CEP-4G (minimal basis and ECP), CEP-31G (split-valence double-zeta and ECP) and CEP-121G (split triple-zeta and ECP). The use of basis set/ECP combinations can reduce the computational cost considerably.
Other ECPs available in Gaussian and in the EMSL basis set library were also tested but did not perform very well compared to CEP.

It was found that very good results can be obtained with these CEP ECP-basis sets. CEP-4G outperforms the all-electron (but minimal) STO-3G basis set for the shielding and CEP-31G performs even better at slightly increased computational cost. The triple-zeta CEP-121G is, however, not an improvement over CEP-31G and it is instead more important to add polarisation functions which we demonstrated by adding an f polarisation on V (from the def2-QZVPP basis set; other polarisation functions had only small effects).

The use of these CEP ECP-basis sets for single-point QM/MM calculations thus seem to be attractive choices to capture quantum mechanical polarisation effects on the chemical shift and EFG tensors, when the QM region needs to be increased beyond the center molecule. Future studies should explore the use of them in more detail, especially for systems with stronger interactions with the environment.

Finally, we note that these convergence studies were performed with standard contracted basis sets which according to the results of Chapter 3.3.2 are not the best choice for the EFG tensor. As we are not comparing to experiment here this should be of little relevance and there is no reason to believe that decontracted basis sets of layers 2 and 3 should be necessary as these layers only serve the purpose of polarising the density around the NMR nucleus.

3.4.3 Results: Comparison with experiment

In our original analysis and comparison of the computed anisotropic chemical shift and EFG data with the experimental solid-state NMR data,[174] we employed normal contracted basis sets and tested functionals that at the time seemed to be the best choices for chemical shifts (B3LYP) and EFG tensors (CAM-B3LYP and LC-ωPBE).

As decontraction of the metal atom basis set now seems to be crucial for converged EFGs and TPSS/revTPSS and their hybrid versions performed better than others in the new analysis in Chapter 3.3.2, we have now performed additional calculations for comparison with experiment. Additionally, the promising BB95 and τ-HCTHh(10%) functionals for chemical shifts (Chapter 3.2) have been tested.

The older results have been summarised in Table 12 which contain single-point calculations of cluster III using the PBE, B3LYP and LC-ωPBE functionals.

Table 12 Calculated chemical shift and quadrupole coupling data for the periodic solid (GIPAW) and for embedded cluster model III (with updated point charges) using different functionals, compared with experimental ^{51}V solid-state NMR data. DFT/QZVPP gas-to-solid shifts (ppm), reduced shielding anisotropies, $\delta_\circ$ (ppm), C_Q values (MHz) and asymmetry parameters, η (unitless).[a]

	GIPAW-PBE	PBE	B3LYP	LC-ωPBE	Expt.
Gas-to-solid shift, $\Delta\delta$	15.0	14.3	16.0	7.3	
$\delta_\circ$	406.3	411.9	414.8	334.0	323.33[b]
$\eta_\circ$	0.013	0.018	0.019	0.012	0.031[b]
C_Q	6.10	6.81	6.82	5.57	±5.7[c,d]
η_Q	0.073	0.069	0.073	0.101	0.09[c]

[a] Basis set specification in layers according to Table 9.
[b] Values from Ref.[186]. Tensor values converted into the Haeberlen-Mehring-Spiess convention.
[c] Values from Ref.[173]
[d] The sign of the quadrupole coupling constant is not obtained from solid-state NMR spectra.

Decontracted basis calculations (only on V) with additional functionals were performed, but on cluster II instead, due to the high computational cost of calculations on cluster III. Results are shown in Table 13. They reveal that decontraction of the def2-QZVPP basis set has only a small effect on the results for this system.

Table 13 Decontracted basis calculated chemical shift and quadrupole coupling data for embedded cluster model II (with updated point charges) using different functionals. Reduced shielding anisotropies, $\delta_\circ$ (ppm), C_Q values (MHz) and asymmetry parameters, η (unitless).[a]

	PBE		B3LYP		LC-ωPBE		TPSS	TPSSh	BB95	τ-HCTHh (10%)	Expt.
	Contr.	Decontr.	Contr.	Decontr.	Contr.	Decontr.	Decontr.	Decontr.	Decontr.	Decontr.	
$\delta_\circ$	408.5	416.9	409.7	418.7	328.9	335.6	392.8	377.2	439.2	372.2	323.33
$\eta_\circ$	0.019	0.020	0.020	0.018	0.013	0.012	0.019	0.018	0.022	0.022	0.031
C_Q	6.83	6.95	6.83	6.61	5.59	5.82	6.94	6.66	7.16	6.95	±5.7
	(6.31)[b]	(6.42)	(6.31)	(6.10)	(5.16)	(5.37)	(6.41)	(6.15)	(6.61)	(6.42)	
η_Q	0.069	0.072	0.072	0.077	0.101	0.101	0.072	0.077	0.071	0.074	0.09

[a] Basis set specification in layers according to Table 9 except a decontracted def2-QZVPP basis set is used on V.
[b] Values in parentheses use Q = -4.8±0.1 fm^2 instead of -5.2 fm^2.

The data in Tables 12 and 13 reveal a slight functional dependency for some of the anisotropic parameters. Curiously, the best agreement with experimental results for both the chemical shift and quadrupole coupling data is obtained with the LC-ωPBE functional, even though that functional did not perform so well in either of the benchmark studies in Chapters 3.2 and 3.3.2. The reduced anisotropy parameter is substantially overestimated by all functionals (except LC-ωPBE) and BB95 and τ-HCTHh(10%) are not much better than others functionals. The chemical shift and EFG asymmetry parameters are not very sensitive to the functional. The NQCC is overestimated by all functionals by 1.0-1.5 MHz except LC-ωPBE (~0.1 MHz deviation). Here it must be noted that our EFG benchmark study revealed that functional errors for ^{51}V NQCC computations (see Table 8) are on the order of 1.0-1.5 MHz so this disagreement with experiment is to be expected. The very good

performance of LC-ωPBE (for $\delta_{\circ}$ and C_Q) may just be a fortuitous result but should be explored further in the future.

Overall, experiment and computation may be associated with a noticeable uncertainty and methodological error, respectively, and hence we can only draw limited conclusions from this comparison. There seems to be a sensitivity of the NQCC and reduced anisotropy, however, with regard to the density functional used, and these effects are even slightly larger than the solid-state effects of this system. Better DFT methods are most likely required for improved agreement with experiment.

Finally, we compare our cluster results with that of a fully periodic calculation using the GIPAW NMR method[117] as implemented in the CASTEP program.[187] Computational details are available in the published study.[174]

As the GIPAW NMR calculations used the same crystal structure (BP86-D optimised), the PBE functional and a rather large planewave basis set (energy cutoff of 58.8 Ry) we can compare the results of our large basis cluster calculations with the periodic ones at a rather similar level of theory. The results are shown in Tables 12 and 13. While the gas-to-solid shift, reduced anisotropy and asymmetry parameters for cluster and periodic results are in very good agreement with each other, the NQCC is rather different (0.7-0.9 MHz). Going from a contracted basis to a decontracted basis (on the metal) in the cluster calculations, the difference between periodic and cluster results is increased. Since the value of the NQCC changes very little between cluster II and cluster III models it seemed unlikely that this difference was due to cluster truncation effects. Separate periodic calculations of a single molecule in a supercell revealed that this difference was either due to different levels of basis set saturation for molecular and periodic calculations or due to a sensitivity with regard to the effective core potential used in the GIPAW calculations. It was found that the NQCC changed when a different ECP was used (See footnote no. 40 in Ref.[174]) so it seems that there is a sensitivity towards ECPs in EFG tensor calculations using the GIPAW approach. However, apart from this uncertainty with the NQCC, the agreement between cluster and periodic approaches for describing solid-state effects is rather satisfying.

Finally, we note the uncertainty associated with the experimental value for the ^{51}V quadrupole moment, cited as -5.2(10) fm^2 in Refs.[144,188] A newer value has recently been suggested on the basis of DFT computations, $|Q|$ = 4.8±0.1 fm^2.[172] Our computed NQCC would decrease by ca. 8%, which would improve agreement with experiment for most functionals as can be seen in the values in parentheses in Table 13.

3.4.4 A study of ^{51}V NMR parameters of $VOCl_3$ in all three phases

The study of solid-state NMR properties of $VOCl_3$ was part of a larger study on environmental and dynamical effects of $VOCl_3$ in the gas, liquid and solid phases.

BP86-D Car-Parrinello MD simulations of $VOCl_3$ in the gas-phase (empty supercell), liquid phase (supercell) and the solid phase (doubled crystal cell as mentioned before) were performed for a few picoseconds of simulation time and snapshots from the trajectories extracted. These simulations were performed by Michael Bühl and the details of the simulations can be found in the published study.[174]

NMR calculations using the B3LYP functional and the AE1 basis set (for all molecules) were performed on the snapshots from each trajectory. For the liquid state trajectory, non-embedded calculations on 15-molecule clusters (i.e. similar to cluster II) were performed while for the solid state trajectory, nonembedded calculations on cluster II were performed.
Using the MD averaged isotropic shielding constants for the three phases, the gas-to-liquid shift and the liquid-to-solid shift could be calculated and were found to be: $\Delta\delta_{g\text{-}l}$ = -13 ppm (liquid more shielded) and $\Delta\delta_{l\text{-}s}$ = -2 ppm (solid more shielded).

As the computed gas-to-liquid shift is rather small (compared to typical DFT errors), this suggests that computational ^{51}V NMR studies can safely employ a single $VOCl_3$ as a reference to compute ^{51}V chemical shifts. It was however, found that the net gas-to-liquid shift could be divided up into a geometric effect and an electronic effect. The geometric effect was found to be shielding with a value of -34 ppm while the electronic effect was deshielding, by +21 ppm. It seems thus crucial for computations of environmental effects on transition metal NMR properties to carefully take both effects into account.

The computed liquid-to-solid shift is in qualitative agreement with the experimental value of $\Delta\delta^{exp}_{l\text{-}s}$ =+6.7. The chemical shift range of ^{51}V spans a few thousand ppm and this agreement can be considered satisfactory.

3.4.5 Summary

We have presented a simple cluster-based approach to capture environmental effects on ^{51}V chemical shift and EFG tensors in solid $VOCl_3$ in combination with periodic DFT optimisations of the crystal structure. We found that both the chemical shift and EFG tensor properties converged rather quickly with increasing cluster size and that the embedded point charges were capable of capturing some of the environmental effects present in a molecular crystal. A QM cluster model consisting of the nearest neighbours of the central molecule is large enough for almost completely converged NMR properties with respect to cluster size and compares well to periodic GIPAW calculations. The environmental effects seen in solid $VOCl_3$ are not overly large, however, and it seems likely that the uncertainty of the DFT approximation that is used for the property calculation is the main hindrance to better agreement with experiment.

3.5 Modelling local geometries and NMR properties of molecular crystals by QM/MM

The study of chemical shift and EFG parameters in solid $VOCl_3$ demonstrated that such embedded cluster calculations (single-point QM/MM) on periodic DFT optimised crystal structures are capable of capturing solid-state effects on NMR properties of a molecule in a crystal (although large QM clusters may be needed). An obvious drawback of the approach, however, is the need for periodic DFT optimisations to get an accurate molecular structure from which to build the cluster model. While inexpensive for the small unit cell of $VOCl_3$, periodic DFT calculations will not always be affordable for crystals with larger unit cells. With impressive progress being made in experimental solid-state NMR spectroscopy techniques, there is a need for theoretical calculations to be able to compute chemical shift and EFG properties for molecular crystals of any size, while reliably capturing the solid-state

effects which will, as previously mentioned, be of a geometric as well as electronic nature. As X-ray structures are typically not accurate enough for such single-point property calculations, geometry optimisations in the solid state will be necessary. Periodic geometry optimisations can become very expensive, however, as the unit cell gets larger. An attractive alternative would be to only relax the geometry of a single molecule within the solid phase as this may very well be the most important geometric effect to account for. This might result in accurate enough metal-ligand bonds of transition metal complexes on which NMR parameters strongly depend.

Typically, QM/MM studies require a full force field for the MM region in order to perform geometry optimisations, which is not a problem when modelling proteins, for example. However, for many inorganic or organometallic molecules, and especially transition metal complexes, force field parameters are not always available. The manual labour involved in deriving such parameters, makes QM/MM seem like an unsuitable approach in many cases. This partly explains the lack of QM/MM studies of molecular crystals with transition metal centers. Recent examples of organic molecular crystal modeling with QM/MM approaches include, e.g. the self-consistent Madelung potential semi-empirical method (SCMP-NDDO)[189], the extended point charge and supermolecule (PCX and SMX) methods[190,191], the hybrid QM/MM many-body interaction model[192], modeling of proton transfer and polarons in crystalline diamino-dinitroethylene[193], modeling of photocrystallographic properties of organic molecules in a crystal[194], and QM/MM molecular dynamics simulations of organic (opto)electronic materials.[195] A perspective article just published, discusses recent advances in molecular crystal modelling with QM/MM-based approaches with an emphasis on lattice energies and lattice parameters.[196]

In this chapter we present a general QM/MM-based method to model solid-state effects on small molecules in a molecular crystal. The method does not require a full force field, yet allows geometry optimisation in the solid state. It requires a starting crystal structure (e.g. from X-ray crystallography) but no system-specific parameters (except for vdW parameters, which here are general), is almost black-box, and can be easily automated. We then test the protocol on selected molecular crystals, both for modelling geometrical effects as well as molecular properties. Most of this work was presented as a publication in *Journal of Chemical Theory and Computation* in 2012.[197]

3.5.1 A general QM/MM protocol for local properties of molecular crystals

The protocol is shown schematically in Figure 9. It is an extension of the single-point QM/MM protocol used in Chapter 3.4 and is described as follows:

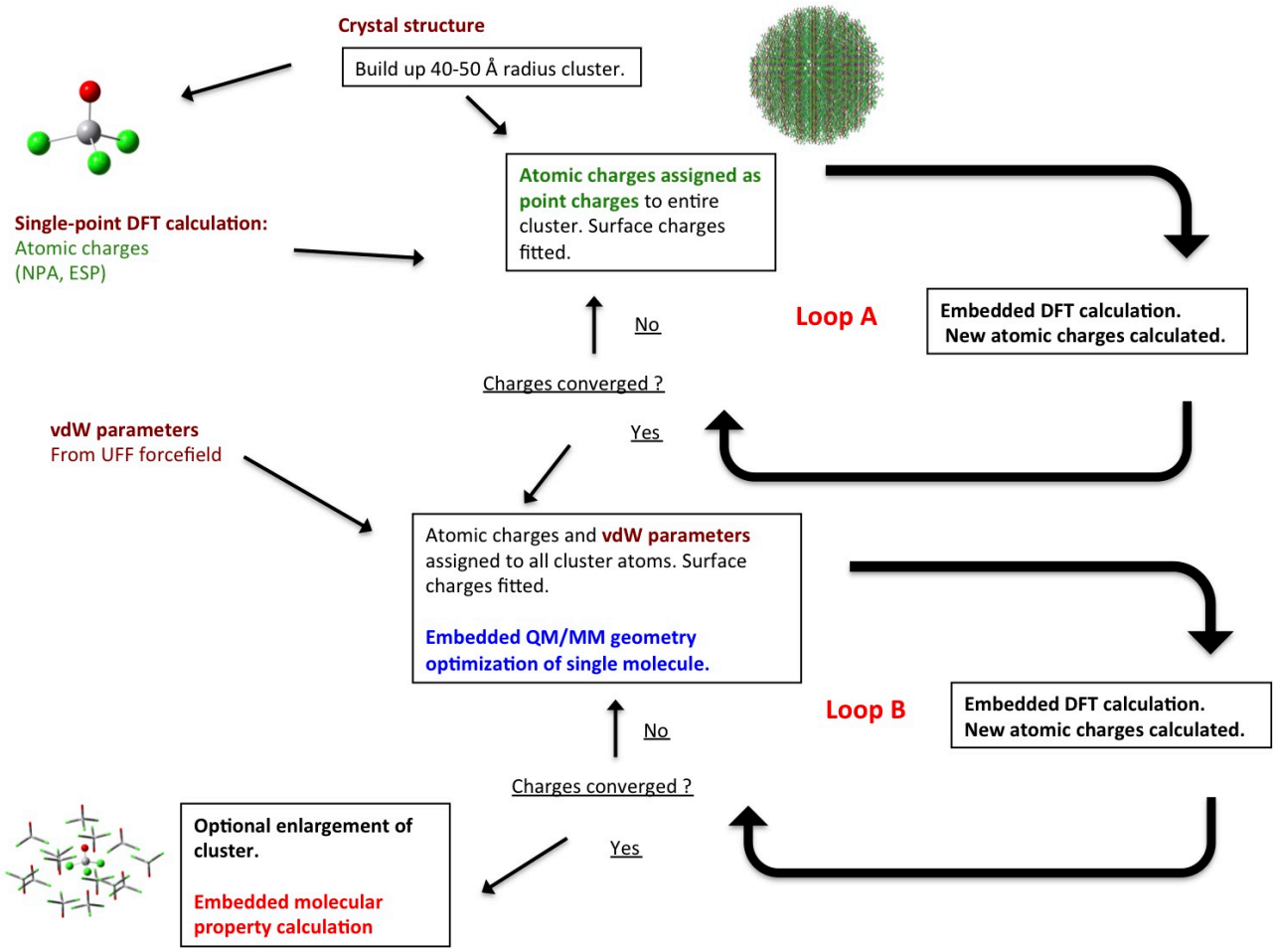


Figure 9 The self-consistent electrostatic embedding QM/MM protocol for geometry optimisation and molecular property calculations of molecular crystals.

1. Crystal structure information from experiment or other source is used to build up a large non-quantum cluster (typically with a radius of 30-50 Å) centered on one molecule. Atomic charges from a single molecule DFT calculation (at the initial geometry) are then assigned to all other equivalent molecules in the cluster. If the cluster consists of multiple molecules, e.g. solvent or counterions, then charges for those need to be assigned as well.

2. Loop A: Additional surface charges are fitted on the outside of the cluster to emulate the electrostatic potential of the infinite periodic system using the new *construct_cluster* procedure introduced in Chemshell 3.4 which is written by Thomas Keal and is based on algorithms in the Construct Program originally developed by Alexey Sokol.[198] A single molecule DFT calculation of the central molecule, now embedded in the surrounding point charges, is carried out. The resulting atomic charges from a population analysis of that calculation are then assigned to all the other equivalent classical molecules. The charge calculations, assigning and surface-charge-fitting is updated self-consistently, typically until the charges change less than 0.0001 e- .

3. Van der Waals (vdW) parameters, taken from the UFF force field,[199] are assigned to all cluster atoms (including the QM atoms) and used to derive Lennard Jones (L-J) potentials for all atom pairs.

4. Loop B: A QM/MM geometry optimisation is performed, keeping only the QM region (the central molecule) active. All classical atoms (each having only an assigned charge and vdW parameter) are kept frozen and interact only statically with the QM region using the previously derived point charges and L-J potentials. When the QM geometry converges, new atomic charges are calculated, which are subsequently assigned to all other cluster atoms. Geometry optimisation and charge iterations on the new updated cluster are performed until the charges are self-consistent and/or no further optimisation steps are required.

5. If resources allow, the QM region can be enlarged at this stage to include nearest neighbours to the central molecule. A geometry optimisation is then performed using the previously determined point charges, with either the nearest neighbouring units fixed or active.

6. Finally an embedded molecular property calculation can be performed, possibly with a larger QM region, on the QM/MM optimised geometry.

The above protocol is straightforward and can easily be automated when the molecular crystal is homogeneous (only one type of molecule) but requires slightly more user input when the crystal includes, e.g. solvent molecules, counterions etc. The protocol allows freedom in choosing the quantum chemical method for geometry optimisation, atomic charge calculation and molecular property calculation (may be different). The protocol would even allow for a molecular dynamics treatment of the QM nuclei (after loop B). The QM method may be a specific DFT functional or even a wavefunction-based method and basis set. The choice will obviously be a balance between computational cost and required accuracy.

How the atomic charges are calculated also allows some flexibility as they can be defined in many different ways. In this work we have trialled ESP charges (i.e. charges that reproduce the electrostatic potential) , NPA charges (from Natural Population Analysis) and Mulliken charges. The ESP charges (without restraints) were calculated by a procedure in Chemshell (here denoted ESP), according to the protocol described in Ref.[77]. The vdW radii used for both vanadium and ruthenium atoms in the ESP calculations (see later) were 3.0 Å and were in addition scaled by 1.5 like all other vdW radii (automatic procedure in Chemshell). We note that as the metal centers tend to be buried inside a complex, the precise choice of vdW radius has little effect on the actual ESP charge. ESP charges according to the Merz-Kollman protocol[76,200] as implemented in Gaussian 09 (denoted ESP-MK) were also trialled for some of the systems, where a vdW radius of 2.0 for vanadium was used. We also note that while we include the surface charge fitting in our protocol, we have not fully explored, how reliable the surface charge fitting is. It depends on several variables such as where the surface charges are chosen to be located and how much of the cluster is used to fit the charges. We did, however, do trial calculations on test systems that suggested that due to the size of the clusters we used, the surface charges had negligible effects on geometries and molecular properties of the central molecule.

The vdW parameters have all been taken from the UFF force field which is available for almost the whole periodic table. The suitability of these parameters for the current purpose is unknown. If one wishes, available system-specific Lennard-Jones parameters or other types of short-range potentials can be chosen instead. For simplicity, L-J potentials with UFF parameters are always used here. We note that

other intermolecular interactions such as charge-transfer are not explicitly modelled by our approach and the exchange-repulsion and dispersion is only modelled classically. Since there is here no quantum mechanical expression for the exchange-repulsion between QM and MM atoms, charge leakage may possibly occur where electron density may "leak" out to the positive point charges, an artifact that would be prevented by exchange-repulsion in a full QM system. This effect is closely related to over-polarisation effects. It has been shown, however, that charge leakage tends not to be a problem in QM/MM calculations as long as no overly diffuse basis sets are used.[201] This problem tends to come to the fore when covalent bonds are cut and saturated with link atoms, which is not necessary in our cases.

We note that the effective fragment potential method[202,203] presents a more elaborate approach to derive the MM parameters from ab initio calculations and includes a multipole expression for the Coulomb interactions and expressions for induction, exchange-repulsion, dispersion and charge-transfer, all parameters derived from ab initio calculations. In our approach, the lack of some of these interactions in the energy expression can be alleviated by increasing the QM region to better account for the short-range interactions, but which obviously can become expensive even for small molecules as many nearest neighbouring molecules might need to be included. Smaller basis sets, possibly with effective core potentials can, however, be used to make such calculations tractable.

The main advantages of our QM/MM approach are that it is in principle applicable to every kind of molecular crystal as long as one is interested in local properties, is not dependent on the availability of force field parameters for each type of molecule and has low computational cost and simplicity (once the protocol is coded). Since only the QM region is being optimised, only external vdW parameters need to be supplied while the point charges are computed from theory. The decision to only optimise the QM region (i.e. keep the entire MM region frozen) is due to the lack of MM parameters for the full system; only non-bonded MM terms are supplied and hence MM coordinates cannot be free variables. It is possible, however, to allow the MM region to be active if one freezes all internal coordinates of the MM molecules. This remains to be explored.

By giving crystal structure information as input and specifying the size of the cluster, QM method, charge model and vdW parameters, steps 1-4 of our protocol can be performed automatically in a single Chemshell input file, as long as the crystal is homogeneous. More complex crystals require slightly more user input. Scripts to carry out the protocol are available in Appendix 1.1 and will be made available in a future version of Chemshell.

In order to put our proposed QM/MM protocol for modelling solid-state effects on molecular crystals to the test, we have identified six different test systems whose geometry or molecular properties are influenced by their solid-state environments. These systems are shown in figure 10. Results for molecules **25**, **26**, **1** and **27** were included in our published paper.[197]

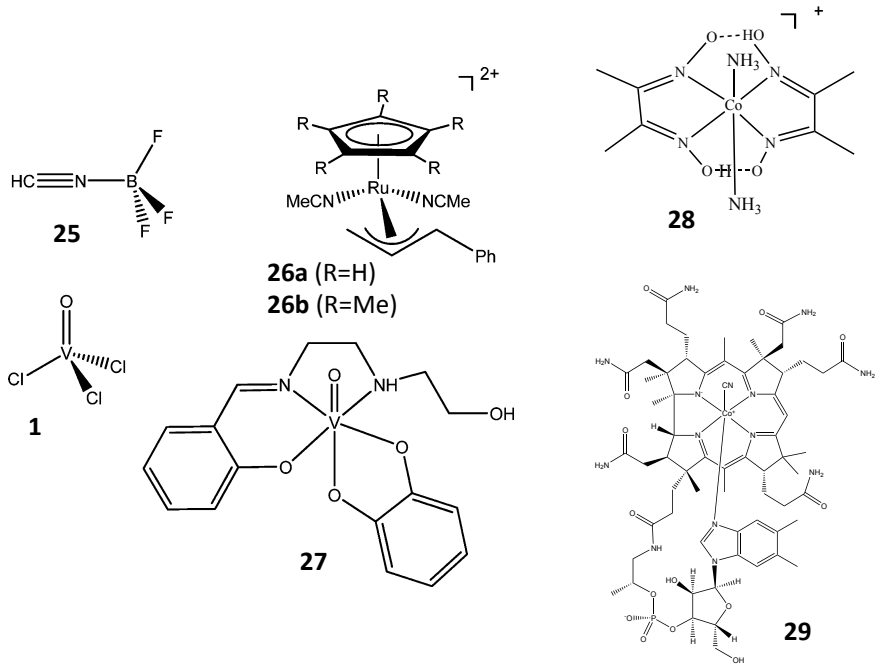


Figure 10 The molecules used to test the QM/MM protocol.

3.5.2 Test case: Modelling the gas-to-solid bond contraction in HCN-BF_3

Molecules can have distinctly different geometries in the gas-phase and the solid phase. The Lewis acid-base complex HCN-BF_3 (**25**) is a good example of such behaviour. According to a microwave study[204] the B-N bond length is 2.473 Å in the gas-phase while the X-ray solid-state structure[205] shows a B-N bond length of 1.638 Å, corresponding to a gas-to-solid bond contraction of 0.835 Å. Similar to the related BH_3-NH_3, [206-208] this large bond contraction has both been attributed to the crystal dipolar field interaction in the solid[209] as well as nearest neighbour short-range interactions.[210]

The gas-phase structure of HCN-BF_3 and other related compounds presents a challenge to theory on its own. Phillips et al.[211] and others[212] have studied HCN-BF_3 and other related molecules and found that these molecules present difficulties for many DFT approximations and wavefunction methods. GGA and meta-GGA functionals often perform very poorly and hybrid functionals like B3LYP still give bond length deviations of 0.06 Å. The hybrid functional B3PW91[16,17,21,213] was found to give a satisfactory equilibrium geometry for HCN-BF_3 with a B-N bond length of 2.465 Å despite not giving a satisfactory binding energy. Equilibrium geometries were also found to be close to vibrationally averaged geometries.

We applied our QM/MM approach to this system in order to see how well such a demanding system can be modelled, especially the gas-to-solid B-N bond contraction. As the molecule is small, this further allows one to expand the cluster to include nearest neighbours (Step 5 of our protocol). As the B3PW91 functional is able to reproduce the gas-phase structure quite well (in contrast to other DFT methods) we chose it for geometry optimisation and atomic charge calculations. It is unclear why this functional should perform so much better than other functionals and is probably due to error cancellations to some extent. A wavefunction method capable of accurately describing medium- to long-range electron correlation effects seems to be necessary for a satisfactory description of this difficult system. We note that it is perfectly possible to use higher level methods in our protocol such as double hybrid

functionals or even coupled cluster theory but that is outside the focus of our study. We have thus for now just used the B3PW91 functional for all gas and solid calculations of HCN-BF_3.

Using the experimental X-ray structure, a cluster with radius 37 Å was built. Mulliken, NPA charges and ESP-MK charges for electrostatic embedding were compared. B3PW91 and the aug-cc-pVTZ basis set[46,48] was used for charge calculations and geometry optimisation and the QM program was Gaussian 09.

The self-consistent loop A (see Figure 9) was iterated 6 times, until the charges changed less than 0.0001 *e*. Loop B was iterated 4 times until the gradient stopped changing between charge iterations.

The QM cluster was then increased to include the 13 nearest neighbouring molecules as shown in Figure 11. The aug-cc-pVTZ basis set was used for the central molecule but the smaller 6-31G* basis set for the neighbours. Two different geometry optimisations were carried out (both using the previously determined self-consistent charges): one with the nearest neighbours frozen at experimental coordinates and the other optimisation allowing all molecules to move. These larger QM optimisations were only performed using NPA and ESP charges.

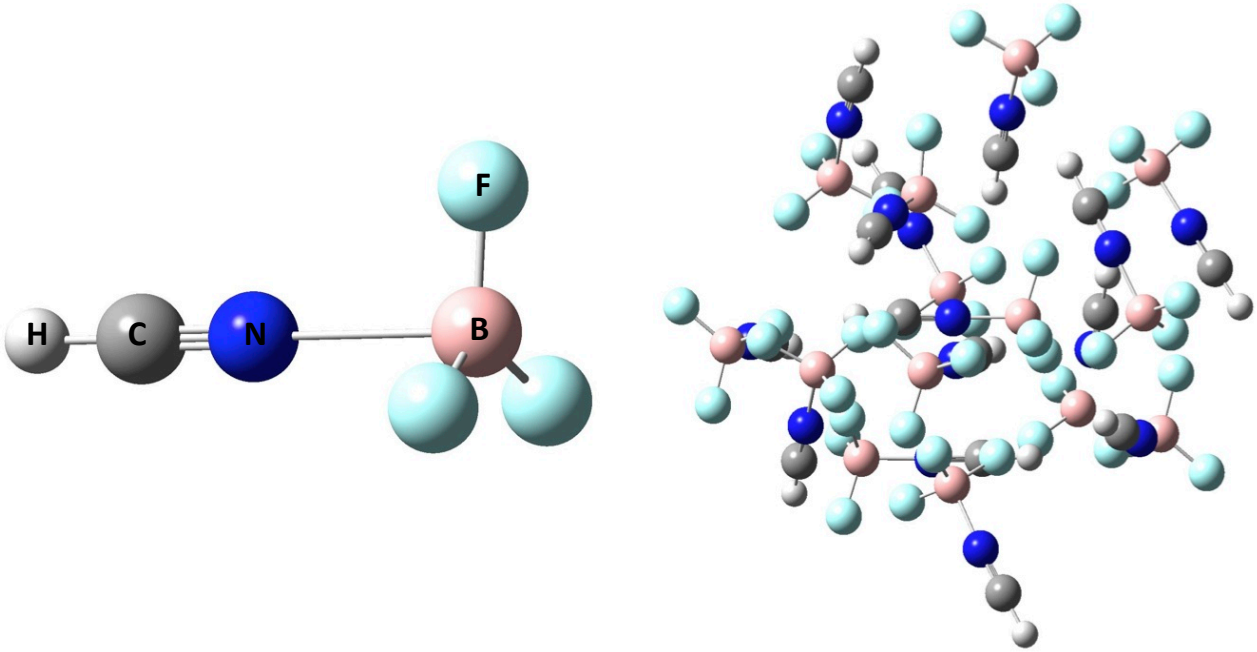


Figure 11 Cluster 1 (left) and Cluster 2 (right) models used in QM/MM calculations of solid HCN-BF_3 (**25**).

Selected bond lengths and bond angles for the central molecule from the different geometry optimisations are shown in Table 14. The B3PW91/aug-cc-pVTZ gas-phase structure shows acceptable agreement compared to the experimental gas-phase structure as deduced from the B-N bond length and N-B-F angle. The solid-state QM/MM optimisations using only a single molecule and with three different charge models for the environment reveal that the B-N bond contraction is very well reproduced using NPA and ESP charges but less so using Mulliken charges. This is not surprising as Mulliken charges are known to be strongly basis set dependent and often rather unreliable.[74] Increasing the QM cluster from a single molecule to 14 molecules (neighbours frozen) results in rather small differences for both NPA and

ESP charge model systems, suggesting that the QM-MM interaction term (electrostatic embedding and L-J potentials) is capable of describing most of the necessary intermolecular effects that influence the potential energy surface of **25**. The C-N-B angle, which deviates slightly from linearity in the single molecule QM/MM optimisations, is improved when including nearest neighbours. Allowing all 14 molecules to move, only results in small changes in bond lengths and angles of the central molecule.

Table 14 Geometric parameters of HCN-BF_3 (**25**) from B3PW91/aug-cc-pVTZ calculations in the gas-phase and in the QM/MM solid.[a]

	Gas	QM/MM Mulliken Cluster1	QM/MM NPA Cluster1	QM/MM NPA Cluster2 frozen	QM/MM NPA Cluster2 relaxed	QM/MM ESP-MK Cluster1	QM/MM ESP-MK Cluster2 frozen	QM/MM ESP-MK Cluster2 relaxed	Exp. gas[b]	Exp. solid[c]
B-N bond length (Å)	2.465	1.572	1.610	1.618	1.614	1.625	1.616	1.609	2.473	1.639
Gas-solid diff. (Å)	-	0.893	0.855	0.847	0.851	0.840	0.849	0.856	0.834	
N-B-F angle (°)	93.5	107.9	105.9	106.0	106.2	105.4	106.1	106.3	91.5	105.6
Gas-solid diff. (°)	-	-14.4	-12.4	-12.5	-12.7	-11.9	-12.6	-12.8	-14.1	
C-N-B angle (°)	180.0	175.3	175.2	177.8	177.4	175.3	177.8	177.2	180.0	178.5
C-N bond length (Å)	1.143	1.135	1.134	1.135	1.135	1.133	1.135	1.135	n.a.	1.121

[a]See Figure 11 for definition of Cluster 1 and 2. [b]From Ref.[204]. [c]From Ref.[205]

These results are encouraging as they suggest that accurate molecular structures can be obtained in a molecular solid using only electrostatic embedding and classical short-range potentials, even for a molecule that experiences huge geometric differences between the gas and solid.

3.5.3 Test case: Solid-state effects on the weak Ru-C interaction in an Ru(IV) allyl complex

In a recent study by Calhorda et al.[214] the geometry of the Ru(IV) allyl dicationic complex, $[Ru(\eta^5\text{-}Cp)(\eta^3\text{-}CH_2CHCHC_6H_5)(NCCH_3)_2]^{2+}$ (**26a**), was calculated at several DFT levels and compared to the experimental X-ray structure[215] of the analogous $[Ru(\eta^5\text{-}Cp^*)(\eta^3\text{-}CH_2CHCHC_6H_5)(NCCH_3)_2]^{2+}$ (**26b**, Cp* instead of Cp). It was found that all levels of theory gave an unsatisfactory description of the coordination mode of the η^3-allyl unit as evidenced by the 0.2-0.5 Å deviation from the experimental value for the Ru-C3 bond (C3 being the substituted allyl carbon). DFT calculations on 4d transition metal complexes typically give bond length deviations of 0.02-0.03 Å according to recent benchmark studies.[216,217] These large deviations were thus interpreted as a failure of current DFT approximations to describe the Ru allyl binding. PBE0 in particular has been found to give very good geometries for 4d and 5d transition metal complexes, hence its failure to describe the Ru-allyl coordination mode is surprising.

Truhlar and Kulkarni[218] found that simplifying the experimental complex **26b** to **26a** in the gas-phase is not entirely appropriate, and newer density functionals like M06-L[219], M06[24], SOGGA[220] and ωB97X(-D)[221] were found to give Ru-C distances in better

agreement with experiment when calculating the real system **26b**. Nevertheless the Ru-C3 distance was still an outlier for all functionals. Jacobsen[222] recently reassessed the problem and made a case for crystal packing effects being responsible for the deviation from experiment as argued from semi-empirical lattice energy calculations.

As our QM/MM method should be able to account for some of the intermolecular forces responsible for such proposed crystal packing effects, we carried out calculations on **26b**. Using the experimental crystal structure (PF_6^- counterions and co-crystallised acetone), a cluster with radius 42 Å was built. The initial QM region consisted of a single dicationic Ru complex which was later expanded into a dimer (See Figure 12). Charge iterations and geometry optimisations of the QM region were performed at the PBE0 level using the mixed def2-TZVPP/def2-SVP basis set (metal/ligand, with the accompanying ECP on Ru) with and without the DFT-D dispersion correction by Grimme[26], trialling both ESP and NPA charges. Only charges for the Ru complex were iterated, while charges for PF_6^- anions and acetone solvent molecules were kept fixed at gas-phase values from DFT calculations at the same level of theory. Turbomole was used as the QM program for geometry optimisations while the (embedded) charge calculations were performed with Gaussian 09.

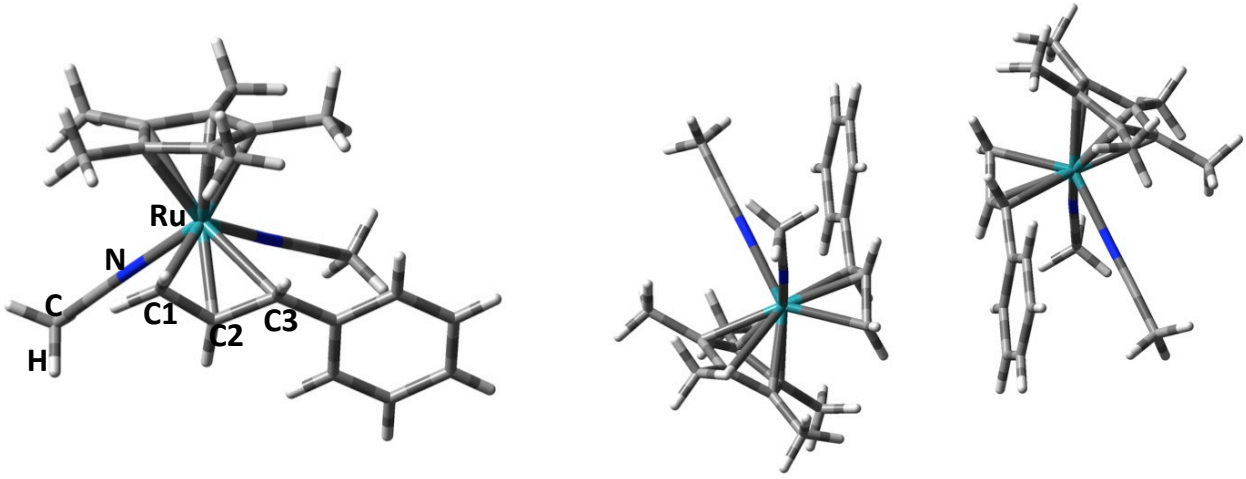


Figure 12 The monomeric (left) and dimeric structures (right) of the ruthenium allyl complex **26b**.

Table 15 Geometric parameters of the Ru-allyl complex, **26b** from PBE0(-D) calculations on monomeric and dimeric units in the gas-phase and in the solid. Bond lengths and deviations (exp. – calc. in parentheses) in Å.

		Gas	Gas	QM/MM ESP	QM/MM NPA	QM/MM ESP	QM/MM NPA
		Monomer	Monomer	Monomer	Monomer	Dimer	Dimer
	Expt.[b]	PBE0	PBE0-D	PBE0-D	PBE0-D	PBE0-D	PBE0-D
Ru-C1	2.181	2.156	2.157	2.157	2.165	2.161	2.160
		(0.025)	(0.024)	(0.024)	(0.016)	(0.020)	(0.021)
Ru-C2	2.189	2.202	2.185	2.167	2.170	2.170	2.168
		(-0.013)	(0.004)	(0.022)	(0.019)	(0.019)	(0.021)
Ru-C3	2.382	2.501	2.426	2.360	2.346	2.354	2.351
		(-0.119)	(-0.044)	(0.022)	(0.036)	(0.028)	(0.031)
Ru-C_{Cp^*}[c]	2.222	2.219	2.209	2.206	2.207	2.206	2.207
		(0.003)	(0.013)	(0.016)	(0.015	(0.016)	(0.015)

[a]See Figure 12 for monomer and dimer structures. [b]From Ref.[214] [c]Average Ru-C distance to Cp* ring.

The results obtained in the gas-phase and in the QM/MM embedded cluster are shown in Table 15. Comparing gas-phase PBE0 and PBE0-D results reveals that intramolecular dispersion has a substantial effect on the Ru-C3 bond by shortening it by 0.075 Å and the deviation between theory and experiment is improved from 0.119 Å to 0.044 Å. However, the environment seems also to have an effect: when the dication is allowed to relax at the same DFT level but in the embedded crystal, there is a further shortening of 0.066 or 0.080 Å (depending on the use of ESP or NPA charges), resulting in even better agreement with experiment for the Ru-C3 bond. The Ru-C3 equilibrium distance from the QM/MM calculations becomes shorter than the experimental value which seems reasonable as equilibrium geometries are usually shorter than thermally averaged structures (such as X-ray structures). Accounting for zero-point and thermal motion tends to increase metal-ligand equilibrium distances by at least a few picometers.[216]

QM/MM geometry optimisation of a dimer (using converged charges from the monomer calculation) was also performed, resulting in slight shortening of the Ru-C3 bond when using ESP charges and slight lengthening when using NPA charges. Our results are thus in agreement with Jacobsen's argument that intermolecular effects in the solid state affect the Ru-allyl bonds, although intramolecular dispersion has an equally large effect, which would also explain the results by Truhlar et al. as the best performing functionals in that study were developed to account for dispersion effects. Ideally, both intramolecular and intermolecular effects should be reliably taken into account in any study of a molecule in the solid state (see e.g. Ref.[223]).

We emphasise that our results are not real predictions of how the molecule behaves in the solid, as experimental crystal structure information is directly used in our protocol (this is, however, usually the case in any kind of modelling of the solid state). One might argue that all we have done, is to put experimental constraints on the system. Nevertheless, we have been able to demonstrate that when a simple account of environmental effects by electrostatic embedding and short-range potentials is made, there is no serious disagreement with experiment and DFT results. And when the cluster is enlarged to a dimer, which introduces more quantum effects and relaxes some of the experimental constraints and reduces the dependence on empirical short-range potentials, very similar results are still obtained. We agree with Jacobsen that for a final unbiased, reliable assessment of this system, full periodic DFT optimisations need to be performed at an appropriate DFT level where both the molecular geometry and lattice constants are allowed to relax. Nevertheless, our QM/MM method seems like a promising tool to capture solid-state effects on medium-sized molecules, for direct comparison with crystal-structure data where strong crystal-packing effects can be expected.

3.5.4 Test case: Solid-state ^{51}V NMR parameters of $VOCl_3$: comparison of periodic and QM/MM geometries

In Chapter 3.4 we looked at anisotropic ^{51}V NMR parameters of $VOCl_3$ (**1**) in the solid-state. Periodic BP86-D calculations with a planewave cutoff of 80 Ry, based on the crystal structure from Troyanov (with fixed experimental lattice parameters) were carried out with subsequent single-point embedded NMR calculations. In order to minimise basis set effects on the periodic geometry the planewave cutoff was now increased to 320 Ry, resulting in rather small geometric changes.

We have now performed electrostatically embedded QM/MM geometry optimisations in order to see how they compare to periodic DFT optimisations. QM/MM geometry optimisations with the BP86-D functional and the large def2-QZVPP basis set (same level for the charge calculations) were carried out on a 48 Å cluster built from the non-optimised Troyanov structure, using three different charge models for point charge embedding: Mulliken, ESP and NPA charges. Gas-phase BP86-D/def2-QZVPP calculations were carried out as well.

Instead of comparing the geometric parameters directly, we assess the anisotropic chemical shift and EFG parameters themselves, evaluated by single-point calculations on non-embedded single $VOCl_3$ molecules, taken from different geometry optimisation protocols. The B3LYP functional was used for the chemical shift and EFG calculations with the large decontracted def2-QZVPP basis set on the vanadium atom and the 6-31G* basis set on O and Cl. Ideally the QM/MM geometry optimisations would perfectly reproduce the results obtained using the periodic BP86-D geometry and hence we can compare different QM/MM protocols.

Table 16 Non-embedded anisotropic chemical shift and EFG parameters (B3LYP/QZVPPdecon/6-31G*) on $VOCl_3$ (**1**) geometries from different sources.[a]

	Periodic Opt.	X-ray[b] Troyanov	X-ray[c] Galy	Gas Opt.	QM/MM Mulliken	QM/MM ESP	QM/MM ESP-MK	QM/MM NPA
σ_{iso}	-2325.36	-2249.51	-2202.44	-2340.87	-2358.37	-2333.36	-2339.82	-2340.21
$\delta_{\circ}$	380.00	378.36	389.90	395.94	341.95	377.31	376.49	377.12
$\eta_{\circ}$	0.016	0.018	0.004	0.000	0.047	0.023	0.021	0.014
C_Q	6.88	6.45	6.33	7.46	7.04	7.19	7.09	6.85
η_Q	0.078	0.036	0.037	0.001	0.016	0.039	0.058	0.120

[a] Molecular geometry optimisations were performed at the BP86-D/def2-QZVPP level. [b] From Ref.[177] . [c] From Ref.[176]

Table 16 reveals that the chemical shift and EFG parameters obtained on the QM/MM optimised geometries with NPA or ESP charges are closer to the periodic results than the results from a gas-phase calculation. The QM/MM Mulliken results are not as good. Results obtained on geometries from the Troyanov and Galy X-ray structures also reveal how sensitive the shielding can be to geometric effects from slightly different experimental setups.

Based on these results and the HCN-BF_3 results it appears that in our QM/MM optimisations Mulliken charges are outperformed by NPA or ESP charges. While **1** does not show large solid-state effects on the geometry (as seen through the NMR parameters) these results are still encouraging and suggest that inexpensive QM/MM optimisations can be good approximations to fully periodic DFT optimisations and are preferable to gas-phase optimisations for NMR calculations.

3.5.5 Test case: The disagreement between computed and observed ^{51}V solid-state NMR parameters of a vanadium catechol complex.

Chatterjee et al.[109] recently published a solid-state NMR study on a series of "noninnocent"[i] vanadium(V) catechol complexes. One of the complexes (**27**) was

[i] A ligand is called "noninnocent" when the oxidation state of the metal, to which it is coordinated, cannot be determined unambiguously. For instance, a catecholate-V(V) complex could also be described with a quinone-V(III) resonance structure.

subjected to DFT calculations of the anisotropic shielding and EFG parameters, which revealed considerable disagreement with experiment. The results were also found to be sensitive to the geometry used and to the basis set used.

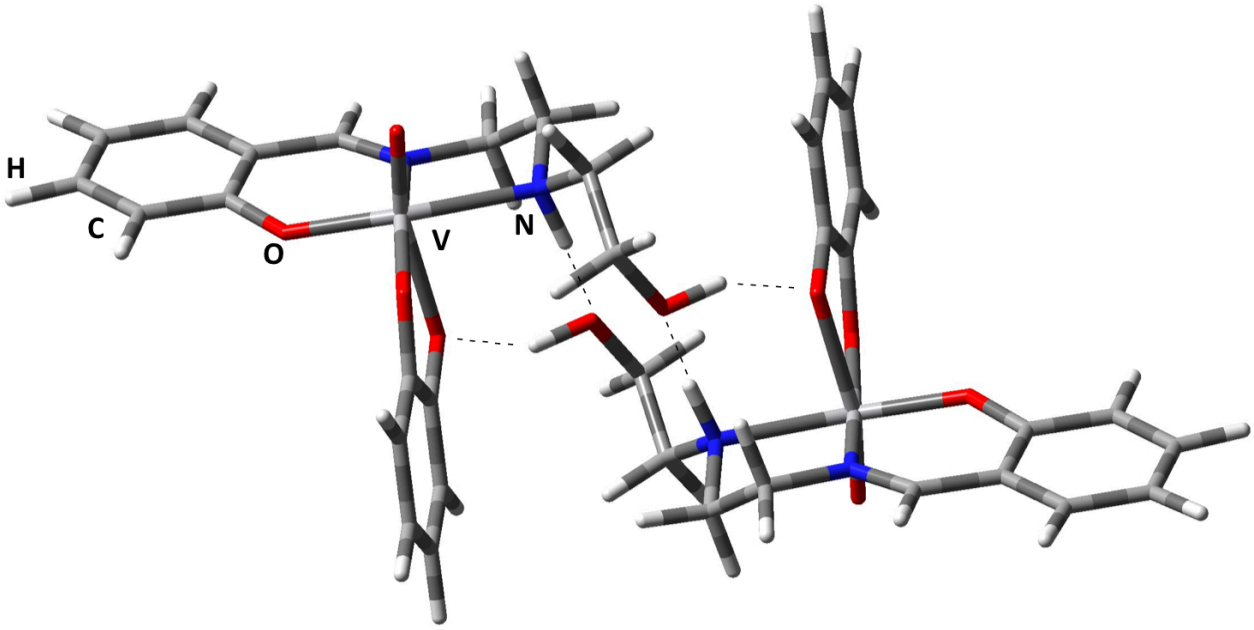


Figure 13 The dimeric structure of the vanadium catechol complex (**27**).

Suspecting a solid-state effect to be the main reason for the disagreement between theory and experiment, we carried out QM/MM geometry optimisations on a 42 Å cluster, using the same X-ray structure[224] as in the work by Chatterjee et al. Inspection of the crystal revealed the presence of two enantiomers of the molecule and hydrogen bonds between the hydroxyl group of one enantiomer with one of the metal-binding oxygen atom of the other enantiomer as well as between the NH and hydroxyl groups (see Figure 13). As these hydrogen bonds may influence the electron density around the V atom and influence the V-O and V-N bond lengths (which the NMR parameters may strongly depend on), we performed calculations with either a monomer or a dimer in the QM region. The latter optimisation was performed after relaxation of the monomer (using the monomer converged charges).

QM/MM geometry optimisations were performed at the BP86-D level using the def2-TZVPP basis set on V, N and O atoms and def2-SVP on C and H. Charge calculations were performed at the same level. Single-point embedded chemical shift and EFG calculations were performed at the B3LYP level on both monomer and dimer structures. In order to minimise basis set effects in the chemical and EFG calculations, the large decontracted def2-QZVPP basis set was used on the V atoms and the 6-31G* basis set on O, N, C, and H atoms, as found beneficial for EFGs (see Chapter 3.3.2). A single $VOCl_3$ molecule, calculated at the same level of theory, was used as reference for evaluating isotropic chemical shifts.

Additionally, periodic DFT optimisations were performed at the BP86-D level in order to gauge the influence of a periodic DFT optimised geometry on the NMR parameters.

The optimised geometry was then subjected to the QM/MM-NPA protocol but without geometry optimisations, in order to get self-consistent charges for the environment. Single-point chemical shift and EFG calculations on monomer and dimer structures were then performed, with and without point charges.

Table 17 Anisotropic chemical shift and EFG parameters[a] of vanadium catechol complex **27**; in normal face: monomer; in italics: dimer.[a]

Geometries:		Gas opt	X-ray	X-ray	X-ray H-opt Non-emb.	QM/MM ESP Non-emb.	QM/MM ESP Emb.	QM/MM NPA Non-emb.	QM/MM NPA Emb.	Periodic geo. Non-emb.	Periodic geo. Emb.
	Expt.[b]		Nonemb.	Emb.							
δ_{iso}	58 ± 5	-53.4	54.4	-66.8		-14.2	-108.8	31.2	-139.0	25.0	
		-246.5	*-76.2*	*-150.3*	*-104.4*	*-169.2*	*-179.3*	*-148.1*	*-182.3*	*-160.1*	-203.3
$\delta_{\perp}$	-243 ± 30	-421.3	-450.3	-389.4		-459.7	-407.2	-527.3	-402.3	459.0	
		-427.8	*-418.5*	*-401.7*	*-405.5*	*-397.8*	*-358.6*	*-447.8*	*-351.9*	*-434.8*	-340.2
η_{δ}	0.93 ± 0.05	0.91	0.94	0.82		0.99	0.88	0.99	0.87	0.95	
		0.61	*0.78*	*0.66*	*0.68*	*0.62*	*0.65*	*0.66*	*0.73*	*0.71*	*0.52*
C_Q	± 4.0 ± 0.1	9.30	7.33	7.05		9.35	8.97	7.59	6.89	6.52	
		7.25	*7.46*	*7.58*	*7.37*	*6.72*	*6.35*	*5.33*	*-4.62*	*6.00*	*5.58*
η_Q	1.00 ± 0.05	0.58	0.42	0.44		0.43	0.45	0.53	0.58	0.61	
		0.64	*0.38*	*0.41*	*0.31*	*0.57*	*0.64*	*0.79*	*0.99*	*0.63*	*0.75*

[a]Evaluated at the B3LYP/QZVPPdecon/6-31G* level on BP86-D geometries (basis set: def2-TZVPP on V, O and N, def2-SVP on C and H); see Figure 13 for the structure of the dimer. [b]From Ref.[109]

The monomer results in Table 17 (in normal face) reveal similar deviations from experiment for the solid-state NMR parameters as in the DFT calculations from the Chatterjee et al. study. Using a gas-phase optimised geometry results in a too large NQCC and a too low asymmetry parameter, η_Q while the shielding parameters compare well to experiment. We note that ^{51}V C_Q values from DFT calculations typically have errors of ~1.4 MHz according to Table 8 in Chapter 3.3.4 and the deviation between theory and experiment on the order of 5 MHz seems thus uncomfortably large. Using an X-ray geometry results in a C_Q value closer to experiment but the asymmetry parameter gets worse. The calculated isotropic chemical shift is in very good agreement with experiment but including the embedded point charges in the NMR calculations results in worse agreement. QM/MM optimisations of the monomer were performed using both NPA and ESP charges. Curiously there is a considerable difference in the results obtained, with the NPA results having a lower C_Q value. None of the monomer results give satisfactory agreement with experiment for both shielding and EFG parameters, however.

When optimising the dimer structure in the gas-phase, only moderate improvement for the EFG value is obtained and the shielding gets worse (see dimer results in italics in Table 17). Single-point energies on a dimer using the X-ray structure also do not show much improvement compared to monomer results and optimising the hydrogen positions (denoted H-opt in Table 17) also does not improve the numbers. However, when a dimer structure is QM/MM optimised using the previously determined self-consistent charges from the monomer calculations, the EFG parameters are in much better agreement with experiment, especially when NPA charges are used rather than ESP charges. The 0.61 MHz deviation in the quadrupole coupling constant is comfortably close to the expected functional error of ~1 MHz for ^{51}V quadrupole couplings.

When using the periodic DFT geometry instead of the QM/MM optimised geometry, the EFG tensor is also improved as compared to using the X-ray geometry (although the NQCC is of the opposite sign compared to the QM/MM-NPA optimisations). The

results using QM/MM optimisations are closer to experiment, however. It is not entirely clear why that is so. The QM/MM optimisations are all-electron calculations (as opposed to the periodic calculations that used pseudopotentials) and use a rather large basis set on the V, N and O atoms which should result in accurate metal-ligand bonds.

The shielding parameters from the QM/MM setup, δ_{iso} in particular, are not in very good agreement with experiment. However, we attribute this to a failure of the DFT functional to describe strongly deshielded vanadium complexes[ii] as discussed in detail by Geethalakshmi et al.[225] In that study, an isotropic chemical shift of -5 ppm for the same compound was calculated in the gas-phase at the B3LYP level with a different basis set and compared to the +220 ppm experimental chemical shift in solution. This deviation is significantly more than the typical ~100 ppm deviation which is typically found for ^{51}V chemical shifts as discussed in Chapter 3.2. Our gas-phase B3LYP determined isotropic chemical shift is -53 ppm. The experimental shift in δ_{iso} in going from solution to the solid is -163 ppm. Our calculated gas to solid (QM/MM dimer with NPA charges) shift is -129 ppm (-208 ppm using the periodic geometry), hence suggesting that the environmental effect on the shielding tensor is at least partly being modelled correctly. It seems likely that the DFT functional error on the shielding tensor is dominating the errors in the absolute values from our calculations, preventing quantitative agreement with experiment.

Our results show that accurate solid-state structures are crucial for accurate EFG parameters but also that specific hydrogen-bond interactions in the solid state may have to be included at the QM level for a correct description of these kinds of systems. Using NPA charges in our QM/MM protocol here results in much better agreement with experiment than using ESP charges. It is unclear why that is the case, but for now we are recommending the use of NPA charges in general. Overall, our QM/MM protocol presents itself as a much better starting point for solid-state NMR calculations of molecules than using either gas-phase optimised structures or X-ray structures.

3.5.6 Test case: ^{59}Co solid-state NMR properties of a cobaloxime complex

^{59}Co was one the first nuclides to be measured using NMR spectroscopy.[226] It has one of the largest chemical shift ranges of all nuclides, spanning 20 000 ppm and generally favourable NMR properties. It additionally has a rather large quadrupole moment making the ^{59}Co chemical shift and EFG tensors excellent probes for electronic structure and molecular geometries of cobalt complexes. Computational ^{59}Co NMR spectroscopy has been studied by Bühl and coworkers.[105,227] B3LYP calculations can be recommended (mean absolute errors of 500-760 ppm for complexes from almost the entire chemical shift range of ^{59}Co), and in some cases accounting for solvation and zero-point and thermal effects is necessary. ^{59}Co chemical shifts remain a challenge, however, and better functionals are most likely needed for improved predictions.

Cobaloxime complexes are popular model compounds of the biomolecule vitamin B_{12} where the dimethylglyoxime (DMG) ligands mimic the complex corrin system of vitamin B_{12} and other cobalamines.

[ii] Compared to "normal" V(V) complexes with isotropic shifts around ca. -500 ppm.

A combined ^{59}Co solid-state and solution-state NMR study of selected cobalamines and cobaloximes with differing axial ligands afforded a set of anisotropic chemical shift and quadrupole coupling data.[228] When the cobaloxime solid-state data are compared to solution data for the isotropic chemical shifts, only small liquid-to-solid shifts are noticeable, 0-70 ppm. As the ^{59}Co chemical shift range is 20000 ppm this suggests very small environmental effects on the chemical shift tensor.

Modelling the anisotropic parameters, however, is a challenge on its own and we were interested in seeing how simple gas-phase calculations of the anisotropic parameters for the cobaloxime compounds would compare to the experimental solid-state NMR data. Initial test calculations revealed unusually large deviations between gas-phase computational results (B3LYP) for the $[CoDMG_2(NH_3)_2]^+$ complex (**28**) and the experimental $[CoDMG_2(NH_3)_2]Cl$ results. While the isotropic chemical shift was in good agreement (~300ppm), the reduced anisotropy deviated by 1000 ppm and the quadrupole coupling constant was off by ~50 MHz. Although the anisotropy, not benefitting from as much error cancellation is more challenging to compute than isotropic chemical shifts, a 1000 ppm error is massive. The quadrupole coupling constant deviation of 50 MHz is also surprising since an 11 MHz B3LYP error is expected based on the benchmark study in Chapter 3.3.

These deviations thus pointed towards significant solid-state effects on the chemical shift and EFG tensors in the $[CoDMG_2(NH_3)_2]Cl$ crystal. We were thus curious to see if our QM/MM protocol would be capable of accounting for these effects. Unfortunately, a crystal structure was only available with Br^- as the counterion instead of Cl^- . While Cl^- and Br^- would be modelled identically in our QM/MM scheme (as a fixed point charge with value -1), the different counterion could mean that the crystal structures are slightly different as the ionic radii are different. While the molecular structure of the complex might not change very much, the unit cell volume could be different. Nonetheless we decided to proceed using the $[CoDMG_2(NH_3)_2]Br$ structure.

A classical cluster with radius ~37 Å was built and subjected to the QM/MM protocol. BP86-D calculations with the mixed def2-TZVPP/def2-SVP basis set were used for both charge calculations and geometry optimisations. NMR calculations were performed using a decontracted def2-QZVPP basis set on Co and the 6-31G* basis set on ligand atoms and using three different functionals: TPSS, TPSSh and B3LYP. The isotropic shieldings were converted into chemical shifts by referencing to the $Co(CN)_6^{3-}$ anion.

As the experimental reference is the 1M aqueous $K_3[Co(CN)_6]$ and it has been shown by CPMD simulations that there is a substantial gas-to-liquid shift[105] we here add a $(\delta^{CPMD}_{ave,\,liq} - \delta^{CPMD}_{ave,gas}) = +1205$ ppm gas-to-liquid correction for the standard to all our isotropic chemical shifts. This correction is evaluated as the difference between the average isotropic shielding at the B3LYP level using geometries from BP86/CPMD/D_2O MD simulations ($Co(CN)_6^{3-}$ in periodic box of D_2O liquid) and the average isotropic shielding at the B3LYP level using geometries from BP86 MD simulation in the gas phase.[105,227] We note, however, that geometries and shieldings used to calculate this correction were evaluated using different basis sets than used in our work. It is not known how reliable this correction is but we use it here as it is more likely to improve predictions rather than introduce artifacts.

Table 18 shows the results of chemical shift and EFG calculations of the $[CoDMG_2(NH_3)_2]^+$ cation in the gas-phase and the QM/MM solid with and without embedded point charges. Results for the gas-optimised cation show that the isotropic chemical shift is in good agreement with experiment (after the gas-liquid shift for the standard is taken into account), while the computed reduced anisotropy and the NQCC are in strong disagreement with experiment. Using the X-ray geometry instead, results in even worse agreement, particularly for the NQCC. Interestingly, however, embedding the cation in self-consistent NPA charges results in a dramatic improvement in the reduced anisotropy and especially in the quadrupole coupling constant (a shift of 35.2 MHz) with the isotropic shift being barely affected. The chemical shift and EFG tensors of compound **5** are thus influenced unusually strongly by the surrounding environment.

Some properties change unexpectedly upon QM/MM geometry optimisations. A much lower NQCC is predicted (very close to to experiment) but both the reduced anisotropy and the isotropic chemical shift are now severely underestimated.

Table 18 Anisotropic chemical shift and EFG parameters of the cobaloxime complex (**28**) using different computational models.[a]

	Experiment[b]	Cation Gas opt.	Cation //Xray	Embedded cation //Xray	Cation QM/MM opt.	Embedded cation QM/MM opt
δ_{iso}	5320 ± 100	5733	5022	4964	4116	4127
$\delta_\circ$	-1570 ± 50	-2502	-2625	-2022	-1427	-876
$\eta_\circ$	0.2 ± 0.2	0.31	0.05	0.03	0.23	0.46
C_Q	±30.7 ± 0.4	81.4	99	63.2	-56.2	33.6
η_Q	0.6 ± 0.2	0.76	0.53	0.73	0.7	0.94

[a] Evaluated at the B3LYP/QZVPPdecon/6-31G* level on BP86-D geometries (basis set: def2-TZVPP/def2-SVP). +1205 ppm gas-to-liquid shift correction added to isotropic chemical shifts. [b] From Ref.[228]

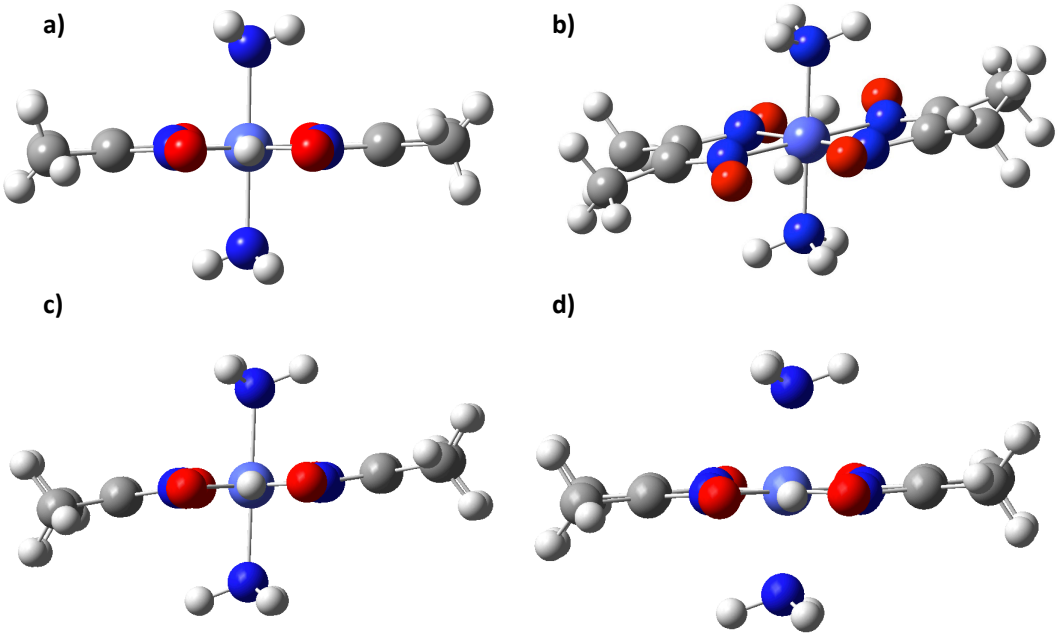


Figure 14 $[CoDMG_2(NH_3)_2]^+$ (**28**) structures from different sources. a) X-ray b) Cation QM/MM optimisation c) Large cluster full optimisation d) Large cluster, partial optimisation.

The reason for these dramatic changes in the tensor properties becomes obvious by comparing the new QM/MM optimised geometry with the X-ray geometry, as illustrated in Figure 14 (a vs. b) and shown by bond lengths in Table 19. Curiously, the QM/MM optimised structure is dramatically distorted when compared to the X-ray structure where the DMG ligand framework deviates significantly from planarity and the calculated Co-N_{ax} bond lengths in Table 19 (see "Small cluster QM/MM Opt") are significantly shorter than both the X-ray and the gas structure.

Table 19. Co-N bond lengths (Å) of the different geometry optimisations of **28** compared to the X-ray structure.

Bond lengths	X-ray	Gas	Small cluster QM/MM Opt	Large cluster QM/MM Opt: Partial	Large cluster QM/MM Opt: Full	Periodic
Co-N_{ax}	1.96	1.98	1.90	1.99	1.97	1.99
Co-N_{eq}	1.89	1.92	1.90	1.91	1.90	1.92

As the [$CoDMG_2(NH_3)_2$]Br X-ray structure shows an almost planar DMG framework and the gas-phase optimised structure is both planar and has Co-N bond lengths closer to the X-ray structure, this geometry distortion would appear to be an artifact of the QM/MM protocol we employ. The most likely explanation would be overpolarisation of the QM region (specifically the ammine groups) by nearby point charges (mimicking hydrogen-bond accepting oxygen atoms), leading to shorter Co-N bond lengths that distort the whole complex or the inability of the L-J parameters to describe the intermolecular interactions between QM and MM regions sufficiently well. It is hard to distinguish between the two effects but it is possible to test if either of these effects are the reason, by increasing the QM region.

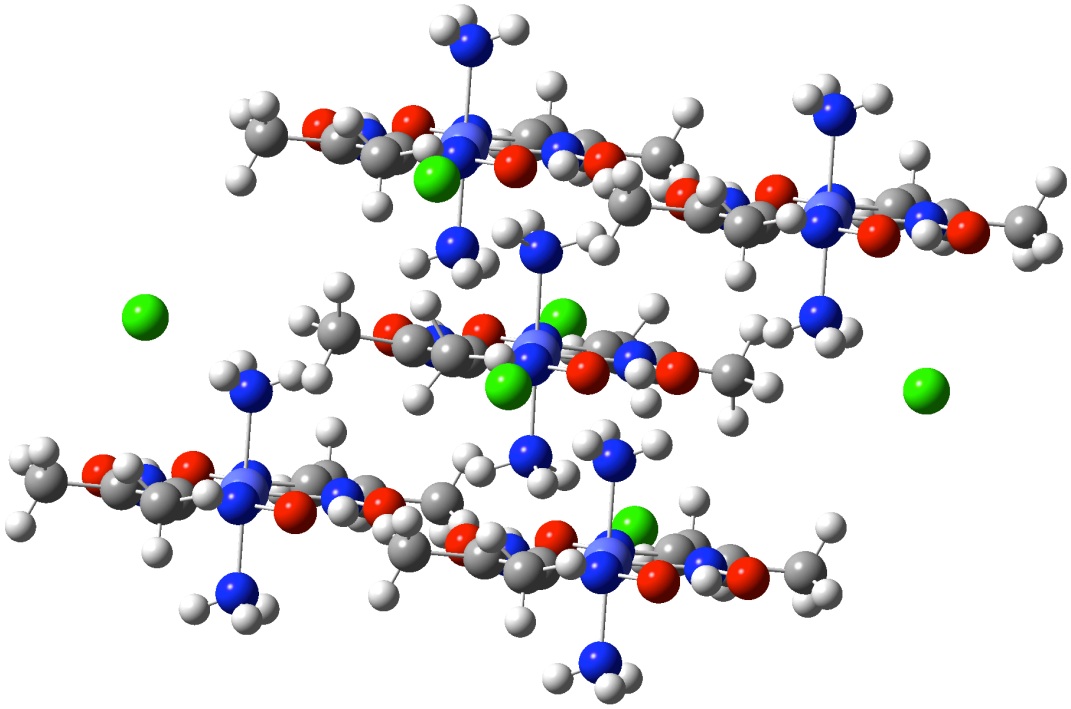


Figure 15 The large cluster of **28** that was calculated.

QM/MM geometry optimisations (using already converged charges) employing a larger -1 charged 5-unit cobaloxime cluster were thus performed. Two different geometry optimisations were tried: one in which the full cluster was optimised and another in which only the central unit was optimised while all other atoms were fixed. The structures from these optimisations can be seen in Figure 14 (c and d). Curiously, optimising the full cluster still results in distortion (of all units), although not as strong as before, while optimising only the central unit (keeping the others fixed but still in the QM region) leads to a structure that looks less distorted. The Co-N bond lengths in Table 19 reveal that the Co-N_{ax} bond lengths for the partially optimised structure are much longer now (presumably preventing most the distortion) but are slightly larger than the X-ray structure. The fully optimised structure on the other hand has bond lengths much closer to the X-ray structure but all units are quite distorted. These optimisations thus demonstrate the sensitivity of the central molecule to neighbouring interactions (specifically the Co-N_{ax} bonds) and suggest the QM-MM interaction term to be at least partly to blame. The reason for why the full cluster optimisation results in distortion might be due to a domino effect where the units near the QM-MM boundary are distorted by point charges which results in distortion of the other units, including the central one.

The NMR results for these large cluster optimisations are shown in Table 20. The results for the fully optimised structure show similar effects due to distortion as the results for the single QM/MM optimised cation before. However, the results for the partially optimised structure (which has almost almost planar DMG ligands) are in remarkable agreement with experiment for essentially all solid-state NMR parameters. The NQCC is slightly larger than experiment, yet an ~8 MHz deviation is to be expected since DFT errors for NQCC can be even larger according to our EFG benchmarking study (see Table 8).

Table 20 Anisotropic chemical shift and EFG parameters of the cobaloxime **28** using the large cluster models.[a]

	Experiment	Large cluster Opt: partial	Emb. Large cluster Opt: partial	Large cluster Opt: Full	Emb. Large cluster Opt: Full
δ_{iso}	5320 ± 100	5776.5	5689.7	5153.1	5070.6
$\delta_{\circ}$	-1570 ± 50	-1680.0	-1629.4	1237.3	-1171.7
$\eta_{\circ}$	0.2 ± 0.2	0.32	0.29	0.45	0.40
C_Q	±30.7 ± 0.4	-36.4	-38.5	25.7	27.4
η_Q	0.6 ± 0.2	0.58	0.58	0.96	0.99

[a] Evaluated at the B3LYP/QZVPPdecon/6-31G* level on BP86-D geometries (basis set: def2-TZVPP/def2-SVP). +1205 ppm gas-to-liquid shift correction added to isotropic shifts.

Additionally, in order to get a more reliable geometry for NMR calculations, we carried out periodic DFT optimisations (using the CPMD code) at the BP86-D level with a planewave cutoff of 80 Ry starting from the X-ray crystal with Br^- exchanged for Cl^- . This structure is much closer to the original crystal structure going by planarity of the DMG framework and the bond lengths are very similar to the partial QM/MM large cluster optimised structure. This demonstrates that the distorted structure obtained from the small cluster QM/MM optimisations are artifacts involving the QM-MM interaction term and not related to the crystal structure having a different counterion (which was another possibility).

Using the periodic structure, the single-point part of the QM/MM protocol was followed which resulted in self-consistent embedding charges. NMR calculations were

then performed for a single cation and a larger cluster with and without embedded point charges. The results, presented in Table 21 show very good agreement with experiment although the reduced anisotropy is slightly underestimated. Again, dramatic changes in the solid-state NMR parameters upon inclusion of the point charges are seen, an effect that is reduced significantly as the QM region is increased.

Table 21 Anisotropic chemical shift and EFG parameters of the cobaloxime **28** using a periodic-DFT optimised geometry.[a]

	Experiment	Cation	Emb. Cation	Large Cluster	Emb. Large Cluster
δ_{iso}	5320 ± 100	5826.3	5734.6	5821.5	5718.7
$\delta_{\circ}$	-1570 ± 50	-2582.0	-1336.2	-1401.1	-1291.5
$\eta_{\circ}$	0.2 ± 0.2	0.09	0.08	0.23	0.19
C_Q	±30.7 ± 0.4	85.0	-28.8	-27.6	-29.2
η_Q	0.6 ± 0.2	0.63	0.42	0.31	0.36

[a] Evaluated at the B3LYP/QZVPPdecon/6-31G* level. +1205 ppm gas-to-liquid shift correction added to isotropic shifts.

We note that the solid-state NMR parameters obtained from these computations are strongly dependent on the DFT method used. We chose the B3LYP functional for most of these calculations as it has been used for ^{59}Co NMR parameters in many other studies but we also wanted to test the TPSS and TPSSh functionals, especially as they were among the best performers in the EFG study in Chapter 3.3. We thus repeated the NMR calculations on the periodic DFT geometry with the TPSS and TPSSh functionals. The results in Table 22 reveal that the obtained chemical shift tensor parameters are worse with TPSS and TPSSh than with B3LYP and that the computed NQCC is closer to experiment with TPSSh (and B3LYP) than TPSS. It may thus be that there is a slight advantage of hybrid functionals for transition metal EFG tensors after all, even though TPSS and TPSSh resulted in almost identical mean absolute errors in the EFG benchmarking in Chapter 3.3.

Table 22 Embedded large cluster of **28** (CPMD geometry) with different functionals.[a]

	Experiment	TPSS	TPSSh	B3LYP
δ_{iso}	5320 ± 100	4504.6	4912.5	5718.7
$\delta_{\circ}$	-1570 ± 50	-1033.3	-1119.3	-1291.5
$\eta_{\circ}$	0.2 ± 0.2	0.06	0.11	0.19
C_Q	±30.7 ± 0.4	-41.3	-33.3	-29.2
η_Q	0.6 ± 0.2	0.24	0.21	0.36

[a] Evaluated at the DFT/QZVPPdecon/6-31G* level. +1205 ppm gas-to-liquid shift correction added to isotropic shifts.

Solid **28** is thus a very interesting and difficult system for computational solid-state NMR spectroscopy. The system exhibits an unusually strong electronic effect of the environment on the wavefunction/electron density that affects the chemical shift anisotropy and especially the EFG tensor. Additionally, modelling the system by QM/MM approaches presents unusual difficulties for geometry optimisation that have not been encountered before.

We have demonstrated that these solid-state effects can be accounted for, resulting in much improved agreement with experiment, but at a rather high computational cost (large clusters and periodic DFT optimisations). This system should be very valuable as a test system to further explore QM/MM and other embedding approaches in order

to better account for solid-state effects on molecules. It may well be that slightly larger clusters like the one in Figure 15 may always be required for the NMR calculation (which could be made computationally tractable by using CEP-basis sets as discussed before), however, one would ideally like to be able to perform the geometry optimisation with a small QM region (single molecule in this case). It is clear that a more accurate QM-MM interaction term will be required for this to be possible. Finally, we note that the validity of our gas-to-liquid correction deserves further scrutiny.

3.5.7 Test case: Towards larger crystals, ^{59}Co solid-state NMR properties of vitamin B_{12}

The molecules we have discussed have mainly been of a small or medium size. The advantage of QM/MM approaches for molecular crystals becomes obvious when a rather large molecule needs to be calculated.

The vitamin B_{12} molecule (**29**) consists of 181 atoms which is a system size within the limits of DFT approaches. The crystallographic unit cell of **29** on the other hand consists of 4 vitamin B_{12} molecules and a considerable number of crystallised solvent molecules (propanol and water), adding up to 1013 atoms and a total volume of 8962 Å^3, according to an X-ray study.[229] Such a large unit cell would be very hard to calculate using periodic DFT approaches.

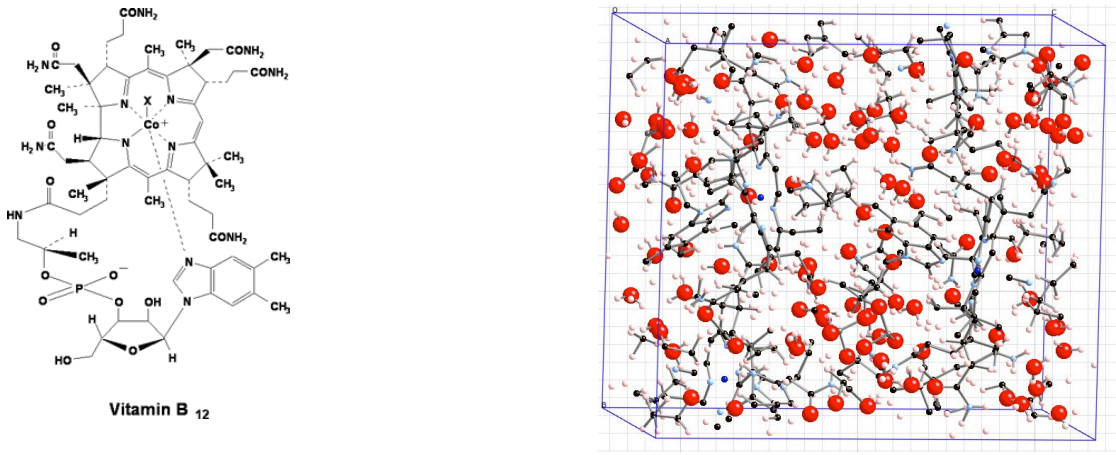


Figure 16 The molecular structure of vitamin B_{12} (**29**) and the unitcell of its crystal structure.

A QM/MM approach to this system is much more manageable as the molecule (or part of the molecule) can be described by quantum mechanics and the rest as static point charges and vdW potentials.

Our interest in this system is fuelled by the availability of high-quality ^{59}Co solid-state NMR data for several cobalamines in both single-crystal[230] and powder form[228,231] and by different crystallisation techniques and differing amount and type of solvent, as seen in Table 23. The agreement in anisotropic chemical shift and EFG data for different experimental conditions (main outlier is the "as purchased sample" which may contain different polymorphs) suggest environmental effects as such not to have a large influence on the tensors but this can be be seen as an advantage as it should offer computational approaches a better chance of reproducing experiment. There is generally good agreement between the single crystal data and the powder data ("as purchased" powder excluded), the chemical shift asymmetry parameter being the

main exception. We decided to use the single crystal data as our experimental reference.

Table 23 Anisotropic solid-state NMR data for vitamin B_{12} (**29**) from different experiments.

	Single crystal (water, "wet")[a]	Powder (as purchased)[b]	Powder (recrystallised)[b]	Powder (water, "wet")[c]	Powder (ethanol,"dry")[c]
δ_{iso}	4549	4800 ± 100	4650 ± 100	4650 ± 100	4650 ± 100
$\delta_{\circ}$	-647 ± 45	-800 ± 60	-650 ± 100	-635 ± 100	-635 ± 100
$\eta_{\circ}$	0.6 ± 0.1	0.2 ± 0.2	0.2 ± 0.2	0.2 ± 0.1	0.2 ± 0.1
C_Q	27.31 ± 0.08	17.7 ± 0.6	26.1 ± 0.4	27.8 ± 0.3	26.1 ± 0.4
η_Q	0.243 ± 0.005	0.2 ± 0.2	0.1 ± 0.2	0.1 ± 0.1	0.1 ± 0.1

[a] From Ref.[230] [b] From Ref.[228] [c] From Ref.[231]

We have thus attempted to reproduce the ^{59}Co solid-state NMR data for this large system using a QM/MM approach. The most recent and most accurate crystal structure was used which contains crystallised water and propanol.[229]

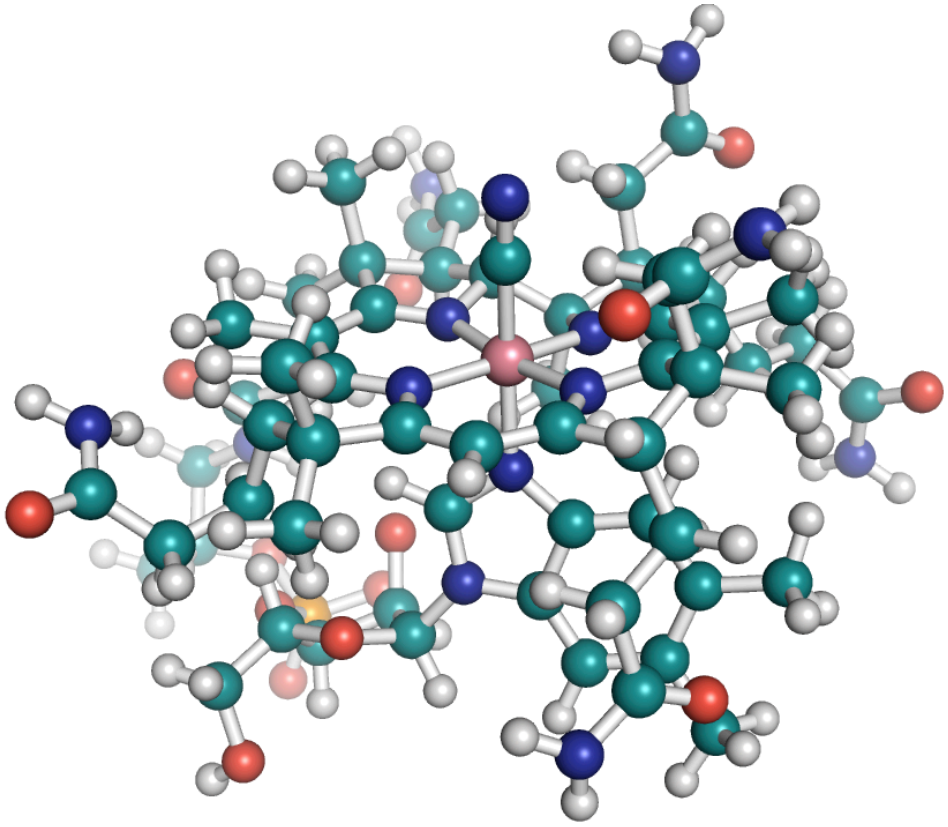


Figure 17 The molecular structure of vitamin B_{12} (**29**).

An older version of our QM/MM approach was used with no iterative updating of point charges. A 42 Å cluster was built up. Point charges of the environment (**29** units, water and propanol) were calculated by separate gas-phase calculations of each molecule at the B3LYP/def2-TZVPP level and using NPA charges. Geometry optimisations were performed at the BP86-D level with the def2-TZVPP basis set on Co, CN ligand, and the Co-bound nitrogens while the def2-SVP basis set was used on on all other atoms. NMR calculations were performed at the B3LYP level with a decontracted def2-QZVPP on Co and 6-31G* on all other atoms.

The results from NMR calculations on a gas-phase optimised structure, X-ray structure and the QM/MM structure are shown in Table 24. The chemical shift tensor seems well described already (although the isotropic shift is overestimated) when using a gas-phase geometry, suggesting that there may be insignificant solid-state effects. The computed NQCC is underestimated and the EFG asymmetry parameter is severely overestimated. We see some changes in the parameters when the X-ray geometry is used instead and when embedding the X-ray geometry with surrounding point charges the EFG tensor changes even more. For the embedded X-ray structure, the NQCC and asymmetry parameter is in rather good agreement with experiment (considering an ~11 MHz DFT error). The isotropic chemical shift is also in good agreement and the reduced anisotropy is in excellent agreement. When a QM/MM optimised geometry is used, the isotropic chemical shift and the reduced anisotropy get worse while the NQCC gets better. The EFG asymmetry parameter is highly sensitive to the surrounding point charges. It increases substantially when the QM/MM geometry is point-charge embedded while it decreases when the X-ray geometry is embedded. It seems thus that vitamin B_{12} does show some solid state effects after all, that present some challenge to first principles calculation of anisotropic NMR parameters.

Table 24 Computed anisotropic chemical shift and EFG parameters of different structural models of vitamin B_{12} (**29**).[a]

	Experiment	Gas opt	X-ray	Emb. X-ray	QM/MM opt	Emb. QM/MM opt
δ_{iso}	4549	5393	4818	4822	4419	4401
$\delta_{\circ}$	-647 ± 45	-663	-585	-653	-774	-730
$\eta_{\circ}$	0.6 ± 0.1	0.45	0.96	0.73	0.66	0.66
C_Q	27.31 ± 0.08	-13.4	10.7	19.0	22.3	-21.2
η_Q	0.243 ± 0.005	0.88	0.52	0.25	0.38	0.94

[a] Evaluated at the B3LYP/QZVPPdecon/6-31G* level. +1205 ppm gas-to-liquid shift correction added to isotropic shifts.

The calculated geometries (bond lengths and angles shown in Table 25) show some changes compared to the X-ray geometry. The Co-N_{imi} (imi: imidazole-ring) bond length changes from 2.09 Å in the gas phase to 2.01 Å in the QM/MM optimised structures whereas the X-ray bond length is 2.05 Å. Additionally, axial ligand atoms form almost linear angles in the X-ray structure while in the QM/MM structure, these angles deviate slightly from linearity. It is thus clear that the environment has an effect on the geometry of **29** which also affects the chemical shift and EFG parameters.

Table 25 Selected geometrical parameters (Å and degrees) for the different geometries of vitamin B_{12}. (**29**).

Bonds and angles	X-ray	Gas opt	QM/MM opt	QM/MM opt + solvent opt
r(Co-N_{imi})	2.05	2.09	2.01	2.00
r(Co-C_{ax})	1.87	1.87	1.85	1.85
r(Co-N_{eq})	1.91	1.92	1.90	1.90
r(C_{ax}-N)	1.16	1.18	1.18	1.17
∠(Co-C-N)	178.7	170.4	172.8	174.7
∠(C-Co-N_{imi})	177.0	177.2	172.9	171.8

In order to explore this further, we decided to increase the QM region and include much of the surrounding solvent (water and propanol) in the NMR calculation as shown in Figure 18. One would not expect the surrounding solvent to affect the NMR parameters much, with the exception of a water molecule that according to X-ray structure seems to be hydrogen bonding to the cyanide ligand. As we have previously shown for the vanadium catechol complex, hydrogen bonding near the NMR nucleus can make a difference for the calculated parameters.

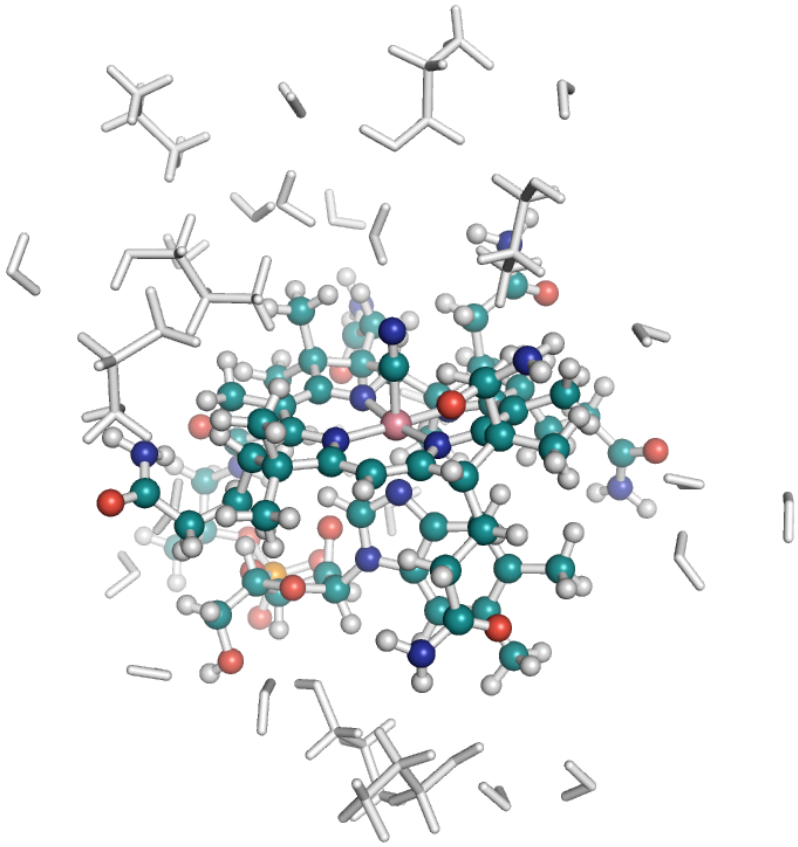


Figure 18 Structure of vitamin B_{12} (**29**) with surrounding solvent in white (propanol and water).

NMR calculations with this extended QM region (with and without geometry optimisations) are shown in Table 26. Geometry optimisations of this larger QM region used the def2-SVP basis set on all solvent, except for the water molecule near the cyano ligand which was described by the def2-TZVPP basis set.

When the solvent is included in the NMR calculations but using the previously optimised (QM/MM) geometry of the cobalamine, the EFG asymmetry parameter stays unusually high while the isotropic chemical shift and reduced anisotropy both increase (in absolute number). When embedding this larger QM region with point charges we see mainly small effects but still some sensitivity of the EFG asymmetry parameter to the environment. When the whole QM region is optimised and the NMR calculations is performed with embedded charges, we see quite a large drop in the isotropic chemical shift and the reduced anisotropy and a smaller drop in the NQCC while the EFG asymmetry stays very much the same. The geometric parameters in Table 25 change only a little upon solvent optimisation.

Table 26 Computed anisotropic chemical shift and EFG parameters of solvent-including models of vitamin B_{12}, **29**.[a]

	Experiment	QM/MM opt + solvent unopt	Emb. QM/MM opt + solvent unopt	QM/MM opt + solvent opt	Emb. QM/MM opt + solvent opt
δ_{iso}	4549	4864	4858	4292	4274
$\delta_{\circ}$	-647 ± 45	-829	-863	-672	-662
$\eta_{\circ}$	0.6 ± 0.1	0.61	0.65	0.60	0.65
C_Q	27.31 ± 0.08	23.5	24.2	22.9	20.6
η_Q	0.243 ± 0.005	0.93	0.78	0.57	0.78

[a] Evaluated at the B3LYP/QZVPPdecon/6-31G* level. +1205 ppm gas-to-liquid shift correction added to isotropic shifts.

Our largest QM/MM model still cannot reproduce the experimental solid-state NMR parameters entirely satisfactorily (η_Q remains too large) but the NQCC and the chemical shift tensor are quite well reproduced. The difference in isotropic chemical shifts, $\Delta\delta$, for the cobaloxime **28** (δ_{exp}=5320 ppm and δ_{calc}=5035 ppm) and **29** (δ_{exp}=4549 ppm and δ_{calc}=4274 ppm) is particularly well reproduced: $\Delta\delta_{exp}$= 771 ppm vs. $\Delta\delta_{calc}$= 761 ppm.

The best agreement with experiment comes from the embedded X-ray geometry where η_Q in particular is in perfect agreement, which may or may not be a coincidence. It is not yet clear what is the best way of describing this system. It may be that a molecular dynamics treatment of the molecule and its surrounding solvent will be necessary, both to check for the sensitivity of the metal-ligand bonds towards dynamical effects (zero-point effects could also be important) which would affect the chemical shift and EFG tensors and for equilibration of the surrounding solvent which may not be in optimal positions as well. Additionally, the uncertainty associated with the DFT method remains (TPSS and TPSSh calculations resulted in very similar NQCC and η_Q parameters).

3.5.8 Summary and future improvements

We have come up with a QM/MM based protocol to model anisotropic chemical shift and EFG properties of molecules in the solid state, capturing the solid-state effects that affect these molecular properties. The protocol allows geometry optimisation in the solid state and is capable of capturing some solid-state geometrical effects as well as direct effects on the chemical shifts and EFG tensors. The protocol can in principle be used for any molecular property that is evaluated by a single-point quantum chemical calculation. Any quantum chemical method that allows embedding in an array of point charges can be used and it will be interesting to see if higher level methods such as coupled cluster theory could be used for molecular property calculations in the solid state for even better accuracy than can be achieved with our density functional theory calculations.
The protocol has been written as a combination of Chemshell Tcl scripts and bash shell scripts. These will be made available in an upcoming version of Chemshell but are also available in Appendix 1.1.

The protocol is quite promising based on the selected examples in Chapter 3.5 but the cobaloxime complex test case shows that there are molecular crystals where the protocol does not work so well and large crystals like vitamin B_{12} are a challenge.

Finally we suggest possible extensions to the protocol that could lead to significant improvements and/or less empiricism. These are:

1. Derive short-range potentials from theory
   Currently the short-range Lennard-Jones potentials used in our protocol uses parameters from the UFF force field. How reliable these parameters are and how well Lennard-Jones potentials can describe the QM-MM interaction is not known. The UFF Lennard-Jones parameters were chosen in our protocol as they are available for almost the whole periodic table. While other vdW parameters or other types of potentials could be chosen for molecular crystals on a case per case basis, depending on availability, an even better option would be if the parameters could be derived directly from information from the calculated electron density of the molecule (like the point charges are) and possibly iterated as well. For the attractive dispersion interaction this would likely require calculation of polarisabilities which might be quite expensive. For the repulsive part we note that an approximate formulation of intermolecular exchange-repulsion has been derived that only uses monomer electron density information.[232]
2. Short-range repulsion at the quantum mechanical level
   Even though short-range repulsion potentials are used in the QM-MM interaction term, there is no repulsion at the quantum mechanical level, meaning that electrons can be artifically attracted to nearby positive point charges and even leak out towards them. In practice, this is unlikely to happen in most of our calculations as very diffuse basis sets are not used so the electrons don't have the flexibility to leak out. Rothlisberger and Sebastiani explored adding a repulsive potential to the Hamiltonian in a planewave DFT code which resulted in electrons feeling repulsion near the MM atoms and found that this slightly sped up the convergence of NMR properties with respect to QM region size.[233] Others have also explored such repulsion potentials in planewave DFT codes.[234] It is likely that charge leakage effects are more prominent in such delocalised basis DFT calculations but others have noted that ignoring exchange-repulsion in QM/MM approaches with atom-centered basis sets, can lead to severe artifacts for excitation energies of a chromophores in a hydrogen-bonding environment.[235] This neglect of exchange-repulsion in our QM/MM calculations may well be part of the reason for the importance of including the hydrogen-bonding partner in the vanadium catechol complex crystal (in Chapter 3.5.5). Including exchange-repulsion in the QM-MM interaction term may thus well be worth exploring further in our QM/MM protocol for molecular crystals.
3. Preventing overpolarisation between QM region and the MM point charges
   While the electrostatic interaction between continuous charge distributions can often be well approximated as a classical point charge interaction at long range (or for QM-MM interaction, the interaction between a continuous charge distribution and a point charge) the same does not apply for the short range. In electrostatic embedding the closest point charges can result in overpolarisation of the electron density.[80] It has been suggested that the point charges be smeared, i.e. multiplied by Gaussian functions for example to prevent overpolarisation. This has been explored as part of more accurately terminating QM-MM covalent boundaries[236-238] but also for better description of hydrogen bonds in aqueous solution etc.[239,240] Smearing of point charges would increase the computational cost somewhat but should not be too

expensive as long as only the MM point charges closest to the QM region are smeared.

4. Beyond electrostatic embedding: frozen density embedding
The approximate interaction between the QM subsystem and the MM subsystem (electrostatic embedding, short-range classical potentials and potential fixes discussed above) becomes much less of a problem the larger the QM region becomes, i.e. the more the nearest environment is described by quantum mechanics. This is already evident in some of test cases we discussed earlier. As the extension of the QM region will increase calculation time considerably this is not always a practical option, even though it may be the most reliable. Alternative embedding approaches have been developed, e.g. frozen density embedding where a traditional QM calculation is embedded by surrounding frozen electron densities.[83-86] For homogenous molecular crystal systems, one could start from a calculated electron density for a monomer and then do an embedded calculation where the calculated monomer density is used as a frozen density for the nearest neighbours and then relax the central molecule density. Such an approach could be very useful as an additional step in our QM/MM protocol to get an even more accurate molecular geometry, assuming that frozen density embedding is not too computationally expensive and sufficiently reliable. Frozen density embedding has even been implemented for molecular property calculations such as chemical shifts.[241] Electron density embedding has also been explored in the QM:QM ONIOM framework.[242]

3.6 Preliminary investigations of the ^{51}V NMR properties of a vanadium dependent chloroperoxidase (VCPO) mutant

We now turn our discussion from NMR properties of small transition metal complexes to NMR properties in a metalloenzyme.

Vanadium is a relatively new entry in the field of bioinorganic chemistry and the functional role of the metal in organisms was unknown for a long time despite the knowledge of its availability in organisms. In fact, only three pages are dedicated to its chemistry in a recently published textbook on bioinorganic chemistry.[243] Despite this, there has been growing interest in the biological role of vanadium, especially after the discovery of the first vanadium dependent enzymes. The bioinorganic chemistry of vanadium has been reviewed many times, [244-246] recently in a very extensive review by Debbie C. Crans et al.[247] and a textbook on bioinorganic vanadium chemistry was recently published.[248]

Well known examples of vanadium in biology include the regulation of phosphate metabolising enzymes where the similarity of the VO_4^{3-} and PO_4^{3-} anions is of key importance, its role as an alternative nitrogenase (a vanadium and iron dependent enzyme expressed when molybdenum is a limiting nutrient), the oxidation of halides by hydrogen peroxide in vanadium dependent haloperoxidase enzymes in marine algae and fungi, the vanadium accumulation as the metal complex amavadine in toadstool mushrooms and the elusive form and role of highly concentrated vanadium

in the blood cells of Mediterranean sea squirts.[247]There has been much interest in the last few years in the potential role of vanadium compounds as insulin mimetics in the treatment of diabetes mellitus[249] and a vanadium compound recently completed a phase IIa clinical trial.[250]

Haloperoxidases are enzymes that catalyse the oxidation of halides (iodide, bromide and chloride) to hypohalous acids in the presence of hydrogen peroxide as the oxidant[251] (Eq. 67).

$$H_2O_2 + H^+ + X^- \rightarrow H_2O + HOX \qquad (67)$$

The hypohalous acid can then halogenate a suitable organic molecule or fulfill some other unknown biological function. One class of haloperoxidases is vanadium dependent (VHPO) where vanadium V(V) acts as the enzyme cofactor, essentially carrying out the catalysis. They have been found in marine algae, fungi and lichens and are believed to be responsible for the majority of halogenated natural products in the marine environment.[252]
Vanadium haloperoxidases are further classified, depending on what halides they are able to oxidise. The first discovered vanadium bromoperoxidase (VBPO) can oxidise bromide and iodide, while vanadium chloroperoxidase (VCPO) can oxidise chloride, bromide and iodide.[251] It has also been shown that these enzymes are capable of catalyzing other reactions, e.g. the oxidation of sulfide to sulfoxide and the oxidation of olefin to epoxide, using hydrogen peroxide as oxygen transfer agent.[253] And as the catalytic efficiency is high, these enzymes have sparked interest for bio-inspired catalyst design.[254,255]

VCPO (which is the focus of our work), from the fungus *Curvularia inaequalis*, is a 67 kDa protein whose structure was determined to 2.03 Å by X-ray crystallography.[256] The X-ray structure shows the vanadium as the central atom of a trigonal bipyramidal complex, with 3 oxygen atoms situated in the equatorial plane, an oxygen atom in one axial position and the nitrogen atom of a histidine residue (His496) in the other axial position.

The cofactor is surrounded by positively charged and hydrophilic residues, obviously hydrogen bonding to the oxygen atoms of the vanadate. The V-N bond length (2.08 Å) is indicative of a coordinate covalent bond and the axial V-O bond (1.88 Å) indicates a hydroxyl/water group while the ~1.60-1.65 Å equatorial V-O bonds suggest unprotonated oxygen atoms. The close distance between the axial O atom and the N atom of a nearby histidine residue (HIS404) suggests a hydrogen bond donation between O4-H and His404-N and other hydrogen bonding interactions can also be deduced. The cofactor with surrounding hydrogen bond donors and acceptors, with atom numbering used later on, is shown in Figure 19.

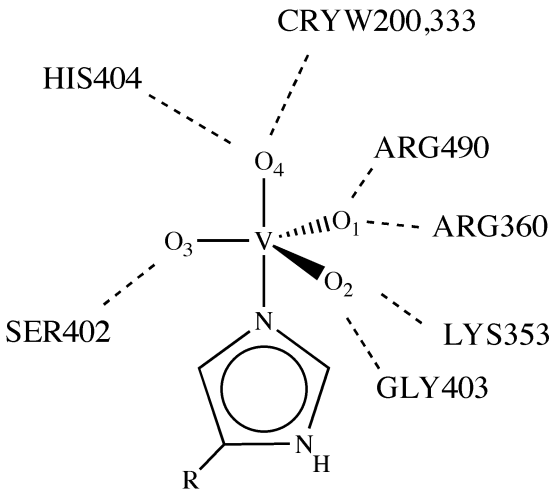


Figure 19. The vanadate cofactor of VCPO and surrounding residues as determined by X-ray crystallography. Protonation state unknown.

As the resolution of the X-ray structure is 2.03 Å, the positions of hydrogen atoms could not be determined. Thus, the actual protonation state of the cofactor cannot be determined from the X-ray structure alone. The V-O bond lengths would suggest that the protonation state is a singly protonated axial O atom, but due to the resolution limitations of the experiment, this cannot be fully determined. EXAFS experiments with much higher resolution (estimated uncertainty on bond lengths 0.02 Å) have also recently been performed that suggests that the cofactor structure may be better described as a distorted square pyramidal with basal V-O bond lengths of 1.54, 1.69 and 1.69 Å (possibly indicating different protonation states), an apical V-O bond length of 1.95 Å and a V-N bond distance of 2.02 Å.[257] The combined structural data from both X-ray and EXAFS is still rather ambiguous with respect to the protonation states of the cofactor and are further complicated by the fact that the experiments were performed at different pH.

Several computational studies have been carried out with the aim of determining the actual protonation state of the resting state of the vanadate cofactor of VCPO as well as the products and intermediates of the full catalytic cycle of the reaction in Eq. 67. Initial studies, which focused more on the catalytic activity, were performed for small 4-coordinate model complexes[258-261] that were not representative of the 5-coordinate cofactor from the X-ray structure. Pecoraro and coworkers studied model complexes using the X-ray coordinates as a starting point and found that doubly or triply protonated vanadate (covalently bound to ammonia) was necessary to stabilise the complex.[262] A singly protonated model, was found to be unstable. In a later study, a doubly protonated model with an axial hydroxo group was concluded to be the most likely resting state.[263] The first QM/MM study, by Carlson, Pecoraro and coworkers,[264] where the protein environment was taken directly into account, confirmed the same doubly protonated cofactor with an axial hydroxo group and one equatorial hydroxo group (O4 and O1) to be a likely resting state but also found a nearly isoenergetic state consisting of both protons on the axial oxygen (i.e. a water ligand). The actual resting state of the enzyme was suggested to be a hybrid of both protonation states. A triply protonated model, consisting of an axial water group and an equatorial hydroxo group was found to be a stable model in the same study, but the authors found it to be unlikely for the resting state. The first step in the catalytic cycle has been suggested to be further protonation of the cofactor, which would make doubly protonated models

more likely than triply protonated models. Raugei and Carloni, employing QM/MM CPMD simulations, found two doubly protonated models to be likely candidates, both models with an axial hydroxo group and the second proton being either on oxygen O1 or O4 (according to Figure 19).[265] An axial water model was not considered.

The above studies, guided mostly by the X-ray structure, have given important insight into possible models of the resting protonation state as well as the catalytic cycle but have nonetheless not given a clearcut answer as to what the actual protonation state of the cofactor is. An important milestone in the study of VCPO was achieved in 2006 with a direct spectroscopic signal from the cofactor. Polenova and coworkers were able to measure a ^{51}V solid-state NMR spectrum of VCPO under magic-angle spinning conditions.[94] In conjunction with the experiment, the NMR properties of several small gas-phase models of possible protonation states were calculated with DFT methods. While the isotropic chemical shift was found to be a poor discriminator of possible models, anisotropic chemical shift and EFG tensors showed significant sensitivity between different protonation states. A doubly protonated model, with an axial hydroxo group, emerged as a likely candidate based on NMR parameters alone but not all possible models found from previous QM/MM studies could be tested as they were unstable as simple gas-phase models.

Bühl and coworkers combined the use of QM/MM calculations, allowing the protein environment to be included, and DFT calculations of chemical shift and EFG tensors of protonation models of VCPO[266]. A range of possible models of the protonation states, from singly protonated to triply protonated, were calculated. Isotropic chemical shifts were again found to be relatively unsensitive with regard to possible protonation models. Parameters from the chemical shift and EFG tensors on the other hand were able to dismiss several models as possible candidates. This included the VOD44 model (doubly protonated axial water group) as the reduced anisotropy had the incorrect sign and the quadrupole coupling constant being two times larger than the experimental value. Three models emerged as the most likely candidates for the protonation state: the doubly protonated models VOD14, VOD34 and the triply protonated VOD144 model. Model labeling refers to the protonation of the oxygens with numbering according to figure 19.

The Bühl group has continued the combined method of QM/MM and anisotropic NMR calculations on the similar enzyme VBPO[267] and on the peroxo forms of VCPO and VBPO[268] for comparison with limited experimental data.
Gascón and coworkers most recently looked at isotropic chemical shifts, geometries and TD-DFT electronic excitation energies of a QM/MM model of VCPO and compared to ^{51}V NMR data, X-ray data and UV-VIS spectra of VCPO.[269] A doubly protonated axial water model was suggested to be the most likely candidate of the resting protonation state, in contrast to the study by Bühl et al. where this candidate was considered unlikely.

There is still considerably uncertainty regarding the cofactor protonation state of this important enzyme and thus additional experimental and computational work is needed. A well known method in biochemistry and molecular biology to investigate local properties of enzymes is to perform site-specific mutations. The crystal structures of such single-amino acid mutations of VCPO have been determined.[270] A colloboration between our group and the the solid-state NMR group of Tatyana Polenova at the University of Delaware (principal investigator of the experimental ^{51}V

NMR VCPO study) has been formed with the aim of exploring the NMR properties of VCPO mutants. Such investigations would give additional insight into the NMR properties of these systems and could help in distinguishing between protonation models for the native enzyme, assuming that the protonation state is the same for mutants and the native enzyme.

Preliminary solid-state NMR data for the R360A mutant of VCPO were recorded by Polenova and coworkers (Polenova, private communication) that showed considerably different anisotropic chemical shift and EFG parameters. In the R360A mutant, the Arg360 residue of the protein has been replaced with alanine. The crystal structure[270] is available from the RCSB Protein Data Bank (PDB ID 1VNF). It shows a cavity due to the missing arginine side chain where a water molecule is now situated, taking over the missing hydrogen bonding interactions to the vanadate cofactor. Overall, there are only very subtle changes in bondlengths of the cofactor for the R360A mutant compared to the native enzyme.

We have performed exploratory QM/MM calculations on the R360A mutant that were compared to the preliminary solid-state NMR data. We learned from our collaborators that the solid-state samples had an excessive chloride ion strength, opening up the possibility of the chloride ions interacting with the active site which is known to inhibit VHPOs, possibly by replacing the OH^-/H_2O axial ligand by Cl^-.[271,272] In view of this we performed calculations on the R360A system with a postulated axial chloride model as well, shown in Figure 20.

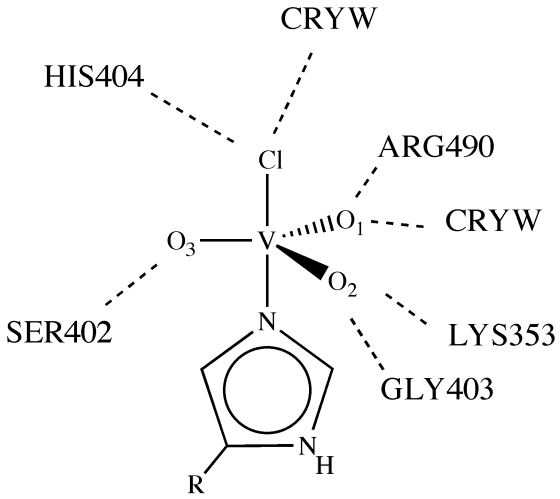


Figure 20. The postulated chloride adduct of the VCPO R360A mutant, where a chloride ion has replaced the axial hydroxo/water ligand.

3.6.1 Computational details

We used approximately the same protocol as has been used for previous QM/MM ^{51}V NMR investigations.[266-268] The following protocol was performed for each model of the protonation state of the R360 mutant and the combined mutant chloride adduct. The X-ray coordinates of the protein (PDB code 1VNF) were imported into the CHARMM program[273] v. 34b2 and the protein was prepared using the CHARMM22 protein force field.[81,82] The protonation states of the histidines of the active sites were assigned as follows: residue His496 (bonded to cofactor) was assigned state HSD,

residue His404 (hydrogen bonded to cofactor) was assigned state HSE (unprotonated N near cofactor; except in the chloride adduct case, see later) and all other histidines assigned state HSP. This differed slightly from previous modelling[266] where His373 was assigned as HSE while His38 and His62 were assigned as HSD but is not expected to make a difference. Neutrality of the whole protein was imposed by protonation of aspartate and glutamate residues at least 5 Å from the active site (aspartates 81, 142, 153, 164, 252, 437, 463, 456 and glutamates 16, 119, 155, 318, 462, 481, 531, 542, 544, 545).

For the cofactor we used the following force field parameters:
The same Lennard-Jones parameters (in kcal/mol) as in the previous study were used:
V atom: ε_X=-0.585 and σ_X =2.15
O atoms: ε_X=-0.1521 and σ_X = 1.77
H atoms: ε_X=-0.0460 and σ_X = 0.2245
where $\varepsilon_{AB}= \sqrt{(\varepsilon_A * \varepsilon_B)}$ and $\sigma_{AB}=\sigma_A+\sigma_B$ holds

Point charges for the vanadate cofactor (for each protonation state) came from BP86/AE1 calculated NPA charges. The vanadium cofactor was kept frozen during all classical force field optimisation and equilibrium steps, thus no bonded parameters were needed. A rigid constraint was applied to the V-N bond to keep the bond in place during equilibration.

The protein was solvated in CHARMM by placing it in a 35 Å water sphere (water molecules described by the TIP3P force field[274]), deleting clashing water residues (those whose atoms came within 2.8 Å distance to a protein atom). Short classical optimisations and MD simulations were run and more water spheres were added in the same place and a loop of deletion and equilibration went on for a few cycles until only relatively few water molecules could be added in each cycle. The total system size was 19418 atoms after solvation. The classical protein preparation finished by a 500 ps classical MD simulation during which all atoms (except the cofactor) within an 20 Å layer from the cofactor were active. Snapshots from the 500 ps trajectory were extracted and used for QM/MM calculations. These snapshots are:
No. 1 at 80 ps, No. 2 at 160 ps, No. 3 at 240 ps, No. 4 at 320 ps, No. 5 at 400 ps and No. 6 at 480 ps.

QM/MM geometry optimisations were performed with Chemshell[89] and the HDLCopt optimiser and HDLC residues.[275] The active QM/MM region was a 10 Å radius sphere around the vanadate, containing ca. 1000 atoms, while the remote parts were kept frozen. Turbomole v. 5.10[276] was used as the QM program and DL_POLY[277] as the MM program using the previously defined force field. BP86/AE1 was used as the QM level of theory. A single QM-region was used which consisted of the vanadate cofactor, and truncated His496, Lys353, Arg360, Arg490, Pro401, Ser402 and Glu403 as well as the 3 nearby crystal water molecules.

We decided to look at the same three protonation models that emerged as likely candidates from the original VCPO QM/MM ^{51}V NMR study (VOD14, VOD34 and VOD144).[266] In addition we looked at the VOD44 model, with an axial water molecule. In some cases, different conformations of protonation model were discovered and were then treated as independent models.

For the chloride adduct models, we used the same force field parameters except for new point charges. The chloride adduct models calculated are: VCL4S1 (singly protonated at O1, charge: -1), VCL4U (unprotonated , charge:-2), VCL4UHP (unprotonated, His404 protonated, charge:-2)

Electrostatically embedded NMR calculations were performed on the QM/MM optimised structures at the B3LYP/AE1 level. We note that these calculations were performed before all other calculations documented in this thesis. We did not employ a computational protocol that would be recommended today. We would recommend reoptimising the QM/MM models at a higher level of theory (larger basis sets on the cofactor) and to use decontracted basis sets for the EFG calculations (to reduce basis set uncertainty) and to test for sensitivity of the results with respect to the density functional. We did, however, test for decontraction of the basis set for a few of the NMR calculations in the Results section which resulted in insignificant changes to the results.

3.6.2 Results

The QM/MM optimised structures of the different R360A models (except chloride adducts) are shown in figure 21.

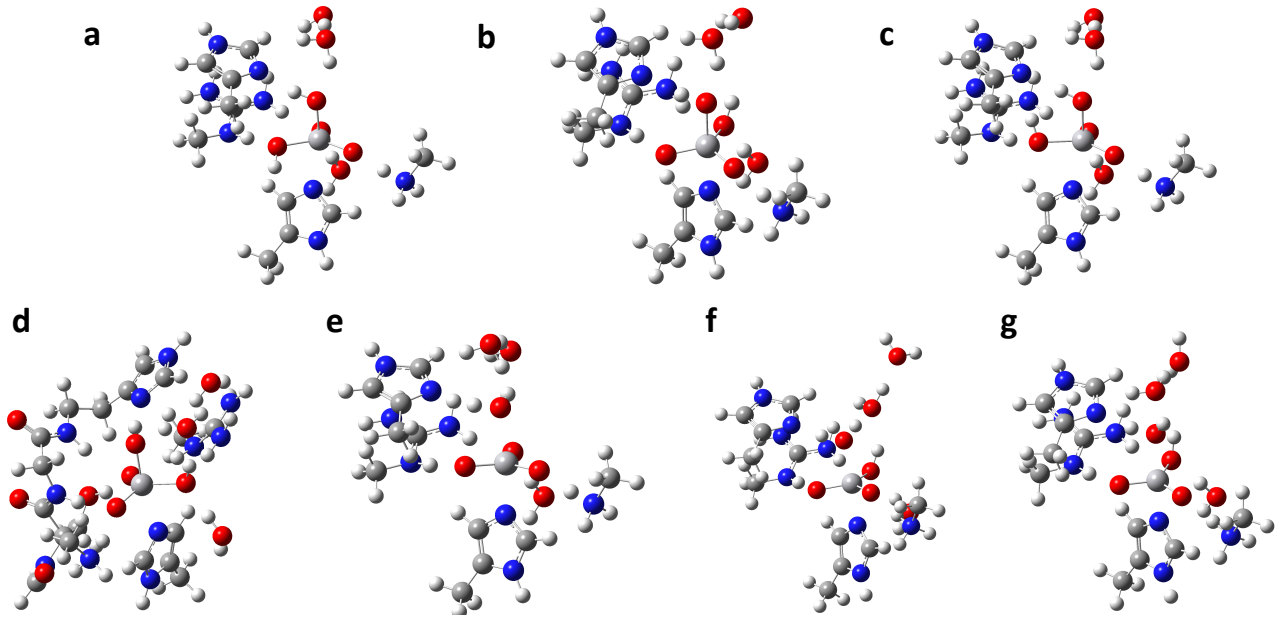


Figure 21 a) VOD34-1: Similar as VOD34-native. O3H pointing towards serine oxygen. **b)** VOD34-2 i and ii: O3H pointing toward His404. i: Ser399-OH pointing away ii: Ser399-OH pointing toward VO3 **c)** VOD14-1: O1H pointing toward O4. **d)** VOD14-2: O1H pointing toward O4. CRYW194 frozen near O4 and O1H during classical MD. **e)** VOD44: axial water turned toward His and V-O1 bond. **f)** VOT144-1: axial water turned toward His and V-O1 bond. **g)** VOT144-2: axial water turned toward His and V-O2 bond.

The preliminary experimental results for the R360A mutant showed only a small change in chemical shift compared to the native enzyme, but a decreased value of $\delta_{\perp}$ and an unusally large NQCC (private communication to Michael Bühl). Due to the

preliminary nature of these experimental results, we cannot show them here. The shifts in these two parameters are substantial and suggest considerable sensitivity in the NMR parameters to geometric changes in the residues around the cofactor (or in the cofactor).

Table 27. Calculated chemical shift and EFG parameters[a] for several protonation models of the R360A mutant. Mean values of the six snapshots and standard deviations.

	VOD34-1	**VOD34-2a**	**VOD34-2b**	**VOD14-1**	**VOD14-2**	**VOD44**	**VOT144-1**	**VOT144-2**
δ_{iso}	-620.07	-595.60	-624.00	-609.66	-580.53	-555.61	-605.43	-632.18
stdev	2.83	9.38	7.57	7.34	5.72	8.79	6.82	11.11
δ_{σ}	-522.08	-616.21	-626.61	-403.92	-381.54	379.72	-611.60	-582.00
stdev	4.86	10.99	8.69	20.41	16.46	104.28	12.39	48.61
η_{σ}	0.254	0.152	0.116	0.563	0.569	0.508	0.382	0.424
std	0.017	0.044	0.036	0.043	0.056	0.092	0.037	0.060
C_Q	-7.70	-9.62	-8.56	-7.04	-8.70	-18.78	-12.02	-9.06
stdev	0.24	0.27	0.30	0.38	0.52	1.50	0.36	2.38
η_Q	0.505	0.704	0.765	0.560	0.459	0.182	0.243	0.540
stdev	0.052	0.096	0.071	0.030	0.039	0.039	0.024	0.267

[a] Evaluated at the B3LYP/AE1 level.

The results for our calculated models in Table 27 show that few of the models predict an increase in NQCC. In fact the only model that predicts a large NQCC is the VOD44 model that in the original study[266] was dismissed as a good model due to the wrongly predicted sign of δ_{σ}. Here we see that the wrong sign for δ_{σ} is again predicted. We looked at the actual shielding tensor for VOD44 and found that the actual tensor simply seems to be incompatible, as large changes would have to occur in order to get the correct sign and magnitude of δ_{σ}. We also note that for one of the snapshots, the axial ligand dissociated, forming a 4-coordinate cofactor, which is at odds with the X-ray structure and is the reason for the rather large standard deviation of the δ_{σ} value. Most models also do not predict a shift of δ_{σ} in the right direction (using values from the original study[266]). The two models based on the VOD14 protonation state, show a shift in δ_{σ} in the right direction. However, no model predicts a substantial increase in the NQCC.

The results for the chloride adduct models are shown in Table 28. The protonated model, VCl4S1, gives a δ_{σ} value compatible with the preliminary experimental observations but the calculated NQCC is still rather low. The unprotonated VCL4U model, however, predicts an even more compatible δ_{σ} value and a rather large NQCC. As the hydrogen bond between the axial O4 group and the His404 N atom is not existent after replacement with chloride (in fact lonepair-lonepair repulsion seems instead likely) we did try a model where the His404 N atom was protonated and thus donating a hydrogen bond to the chloride. The model, VCL4UHP with the cofactor still unprotonated, however, gave the wrong sign for the δ_{σ} parameter and is thus considered unlikely.

One must also bear in mind that no X-ray structure exists for a chlorinated VHPO; our homology model (which may not even be relevant) may reach its limit if chloride

incorporation of the cofactor induces larger structural changes in the protein that may not be adequately captured in our classical equilibration step.

Table 28 The chemical shift and EFG tensors results for the chloride adduct protonation models. Mean values of the six snapshots and standard deviations.

	VCl4S1	**VCl4U**	**VCl4UHP**
δ_{iso}	-506.2	-545.5	-538.8
stdev	10.1	16.2	13.1
δ_{σ}	-453.2	-426.7	394.4
stdev	9.6	55.9	9.6
η_{σ}	0.18	0.64	0.63
std	0.12	0.20	0.13
C_Q	-12.2	-19.5	-19.2
stdev	0.7	0.4	0.7
η_Q	0.35	0.15	0.28
stdev	0.16	0.12	0.03

3.6.3 Summary

To summarise, preliminary investigations into the protonation state of the R360A model were performed by QM/MM calculations of chemical shift and EFG parameters of several different protonation models of the R360A VCPO mutant. Three chloride adduct models were also calculated. Based on these results no standard protonation model seems to fit the preliminary experimental observations sufficiently well, mainly due to the large preliminary experimental NQCC that is generally not predicted by any of the likely computed models. Only the chloride adduct models seem likely to reproduce the preliminary experimental NMR parameters. Our results thus cannot rule out chloride adducts as being responsible for the preliminary R360A ^{51}V NMR spectrum. The uncertainty regarding the experimental results prevents us from making any real assessment but our results can serve as useful computational predictions. Experiments on this and other mutants are being resumed in the Polenova group (private communication to Michael Bühl) but are not sufficiently advanced at the time of writing. It will be interesting to see whether the apparent changes in the ^{51}V NMR characteristics can be confirmed to be inherent to the mutant, or, as suggested by our preliminary results, rather due to the incorporation of chloride into the cofactor.

The combination of experimental and QM/MM computational transition metal NMR spectroscopy seems overall very promising for shedding light on structural

conundrums in metalloenzymes. There is still room for improvement in the computational predictions: better electronic structure methods for more reliable predictions of the anisotropic chemical shift and EFG tensors as well as improvements in computer hardware/DFT algorithms that would enable one to improve phase space sampling by currently too expensive DFT/MM molecular dynamics simulations, instead of the current protocol (QM/MM optimisations of a limited number of snapshots from classical MD simulations).

Chapter 4. Modelling solvation by QM/MM molecular dynamics: the difficult case of the flexible zwitterion 3F-GABA

We now turn from computations of the solid state to the dynamic solution phase where another environmental effect becomes important: solvation.

Solvent effects can play a key role in chemistry, e.g. altering the kinetics of reactions taking place in the solution. Some solvent effects can have an even more dominant role, completely altering the structure and charge distribution of a molecule when going from the gas-phase to solution. A zwitterion is an example of such behaviour, a molecule containing formal positive charges and negative charges in an overall neutral molecule,[278] but is usually only stable in solution or the solid state. Many of the amino acids that make up proteins are zwitterions when found individually in solution. A related molecule is γ-aminobutyric acid, GABA (**30**), technically an amino acid as it contains amino and carboxyl acid groups (but not in the α carbon position like the amino acids that make up proteins). GABA is the main neurotransmitter in the nervous system of mammals where it binds to transmembrane receptors resulting in inhibition of various neural processes. In solution the zwitterionic form is dominant while the form with a protonated carboxylate and an unprotonated amino group (hereafter called the neutral form) is dominant in the gas-phase as shown in Figure 22.

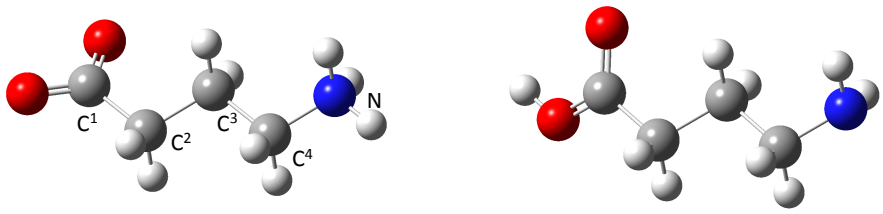


Figure 22 The zwitterionic (left) and neutral form (right) of GABA (**30**). The atom numbering used throughout for GABA and 3F-GABA is shown on the left.

An unsolved problem in neurochemistry is how GABA binds to its receptors, specifically the active binding conformation as well as the structure of the receptors themselves that have yet to be deduced from crystallography studies.[279,280] This question has stimulated synthesis and measurements of various small molecules in order to mimic the structure and function of GABA.[279]

Recently, O'Hagan and coworkers synthesised a fluorinated GABA analogue, 3F-GABA[281] (**31**). The fluorine atom introduces chirality resulting in both *R* and *S* forms of 3F-GABA which have been revealed to bind differently to neuroreceptors.[280] It is believed to have a neutral form in the gas-phase and a zwitterionic form in solution, like GABA. The fluorine atom introduces a conformational bias into the molecule due to the charge-dipole interaction between the C-F bond and the NH_3^+ group and is a convenient spectroscopic probe which can be used to gather information about the conformational properties of the molecule.

The fluorine atom affects the electronic structure and in combination with the chirality results in an interesting conformational energy surface. This surface was

qualitatively described by O'Hagan et al.[280,281] and has been used to explain surprising results from experimental receptor-binding studies where the binding affinity of the enantiomers is found to be different. While the binding conformation of GABA/3F-GABA to its receptors is yet to be deduced, another question is simply, what is the dominant conformation of 3F-GABA/GABA in aqueous solution?

Quantum chemical calculations are a convenient tool to explore conformational behaviour of molecules as they provide access to information often unavailable to experiment.[282] The accuracy of the electronic structure method is often the main unknown in such gas-phase calculations and the environmental effect is often simply ignored with accurate results still being obtained in many conformational studies of molecules. For GABA and 3F-GABA, however, the environmental effect plays a crucial role that cannot be ignored. As the zwitterionic form is simply not stable in vacuum, the environment must be accounted for right from the beginning just to be able to describe the correct structure and charge distribution of the molecule. If a zwitterion of GABA/3F-GABA is optimised in vacuum, spontaneous proton transfer from the amino group to the carboxyl group occurs, revealing the zwitterionic form to be an unstable point on the vacuum potential energy surface.

Theoretical studies of zwitterions have a long history, see e.g. work by Kirkwood where zwitterions are described by Debye-Hückel theory.[283]
More recent studies have focused on whether zwitterions in the gas-phase could exist,[284] how many water molecules are required to stabilise zwitterions[285,286], modelling of the proton transfer between neutral and zwitterionic forms of glycine both in the gas-phase and solution[287-290] to name a few areas.
GABA and its analogues have been studied by Jordan and coworkers[286,291,292] by DFT and MP2 computations of microsolvated molecules and solvated in a continuum, in order to shed light on the conformational properties of GABA in solution.

In the original 3F-GABA study,[281] electronic structure calculations at the B3LYP/6-31+G* level were performed using the PCM continuum solvation model to account for the effect of the aqueous environment. Only the *S* enantiomer of 3F-GABA was considered, as enantiomers are equal in energy and properties in an achiral environment. The study located ten different conformers of zwitterionic 3F-GABA which can be classified into folded (conformers **A**-**D**) and extended (conformers **E**-**J**), shown in Figure 23.

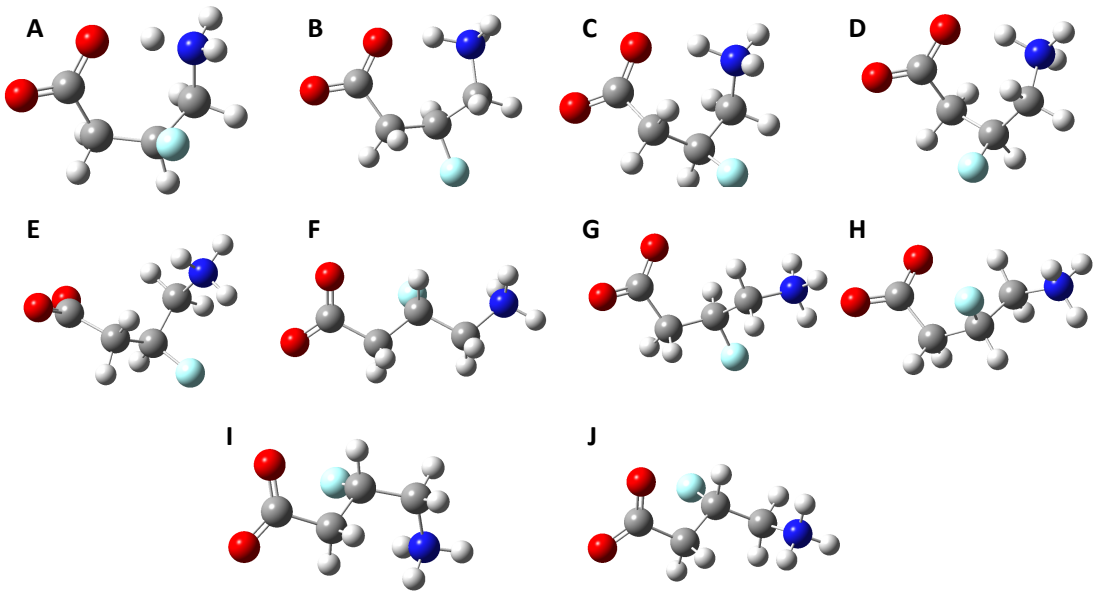


Figure 23 The ten calculated conformers of 3F-GABA (**31**) according to PCM-B3LYP calculations.[281]

The folded conformers contain an intramolecular hydrogen bond between the amino and carboxyl group. Relative energies of the conformers indicated a substantial preference for folded conformers (**A** in particular), presumably due to the stabilising hydrogen-bonding interaction.

In the same study, a detailed analysis of ^{1}H and ^{19}F NMR spectra was performed, resulting in a set of refined $J_{H\text{-}F}$ and $J_{H\text{-}H}$ coupling constants. PCM-B3LYP calculated coupling constants for each conformer were used for iterative refinement simulations of the experimental spectrum which resulted in no single conformer matching well the experimental coupling constant pattern. The closest agreement with the observed spectrum came from data of conformer **F**, however. Additionally, a Karplus relationship that relates ^{3}J coupling constants and dihedral angles indicate that extended conformers like **F** are more likely to contribute to the coupling constant pattern, suggesting that **F** is present and probably dominant in solution. The strong overestimation of some coupling constants ($^{3}J_{F\text{-}H1}$, calc. 51.7 Hz vs. exp. 31.3 Hz), however, was interpreted as additional conformers being populated during experimental conditions and contributing to the NMR spectrum as well.

The PCM-DFT energy analysis and the NMR coupling constant analysis are thus in sharp contrast. The most likely explanation for this would seem to be the approximate environmental effect of the continuum solvation model for the relative energy calculations. Specifically, short-range interactions (such as dispersion, exchange repulsion, charge transfer) may become very important in some systems that continuum solvation models only take into account in an approximate way or not at all (PCM).[3] Benchmarking studies of continuum solvation models for calculation of solvation energies also reveal that ionic solutes are the hardest to calculate accurately.[78] Zwitterions of course share some of the characteristics of ionic solutes and since the amino and carboxyl groups of 3F-GABA are likely to be involved in extensive hydrogen-bonding in aqueous solution, its solvation behaviour may well be rather difficult to describe by continuum solvation models.

The approximate nature of continuum solvation models comes from not representing the solvent as explicit molecules. Thus an obvious improvement in the modelling of solvent effects is the explicit modelling of each solvent molecule. One way of doing this is by microsolvating 3F-GABA by including water molecules around the solute and subject that cluster to DFT calculations . This was attempted by our coworkers in this study, Jie Cao and Tanja van Mourik, who calculated clusters with up to 5 added water molecules, with and without PCM solvation surrounding the cluster. This, however, was found not to substantially change the relative energies between folded and extended conformers (folded conformers still preferred), although just a few water molecules were required to stabilise the zwitterionic form in a vacuum.[293]

The question is whether many more water molecules are required to properly capture the environmental effects that effect the conformational energy surface of 3F-GABA. This becomes expensive to do with a quantum mechanical method and furthermore the increase in degrees of freedom by addition of water molecules start to make traditional geometry optimisations on a potential energy surface troublesome as an extraordinary amount of local minima become available for each solute conformation, each of which would contribute to the free energy. A fully dynamical treatment thus starts to become necessary as more water molecules are described explicitly.

In view of this, we attempted QM/MM molecular dynamics modelling of 3F-GABA where the solute is treated quantum mechanically and the solvent modelled by a parameterised force field. The solute and solvent interact electrostatically (point charges at the MM atoms polarise the electron density of the QM region) and by parameterised vdW potentials. The QM/MM approach allows one to include a large number of (classical) water molecules, thus accounting for both explicit short-range interactions and long-range bulk effects, and the dynamical treatment gives an average over the large number of possible solvent configurations. This, however, comes at the expense of a full quantum mechanical treatment of the short-range interactions of the solute with the nearest water molecules.

Most of this work was published as part of a paper in *Chemistry – A European Journal* in 2012.[293]

4.1 Computational details

The system was prepared classically using the CHARMM program.[273] 3F-GABA was inserted into the center of a pre-equilibrated sphere of TIP3P[274] water molecules (radius 35 Å and 6220 molecules) and subjected to 1 ns classical equilibration at 300 K with the solute frozen and a boundary potential at the surface to keep the solvent in place. Force field parameters for 3F-GABA were derived using glutamate as a template with fluorine parameters obtained from a fluorinated sugar derivative, using the CHARMM27 force field.

QM/MM (Born-Oppenheimer) molecular dynamics simulations were performed using the ChemShell QM/MM program. 3F-GABA was taken as the QM region and all the water molecules as the MM region, QM and MM regions interacting through electrostatic embedding and Lennard-Jones potentials. The inner shell water molecules (3061) were active during the simulation while the outer shell was frozen to prevent evaporation. All molecular dynamics simulations were performed with a time step of 1 fs in an NVT ensemble at 300 K. A Nosé-Hoover chain thermostat[50-53]

with a chain length of 4 and a characteristic time for thermostat coupling of 0.02 ps was used. All hydrogen atoms of 3F-GABA were assigned the mass of deuterium while all TIP3P water molecules had frozen bond lengths and bond angles. This was performed in order to slow down or freeze the X-H vibrations sufficiently for a timestep of 1 fs to be accurate enough for the numerical integration. Turbomole v. 6.2[276] was used as the QM code for DFT/MM simulations using SCF convergence of 10^{-8} au. PM3[36]/MM simulations were carried out using MNDO99 v. 7.1[294] with SCF convergence of 10^{-10} au. The MM code in all QM/MM calculations was DL_POLY.[277]

In order to obtain free energies, constrained MD simulations were carried out using suitable reaction coordinates ξ (dihedral angles) and pointwise thermodynamic integration. The mean forces of constraint were sampled during simulations which were typically 70-150 ps long. A series of statistical tests, checking for trend, normality and serial correlation according to Schiferl and Wallace[295] were performed on the data from all constrained MD simulations. Only equilibrated data (after at least ~40 ps of simulation time) that passed the statistical tests were used to calculate averages and standard deviations. The converged mean forces were integrated over the reaction coordinate affording free energy profiles between reactant *A* and product *B*, according to Eq. 40.

Strictly, the relation in Eq. 40 is only an approximation when using a dihedral angle as a reaction coordinate.[54] We did not take into account corrections for this choice of ξ as they are known to be very small (smaller than the statistical sampling errors).[54] As the MD simulations are performed in the NVT ensemble, Helmholtz free energies ΔA are obtained. Because volume changes in condensed phases are small, these are very close to the corresponding Gibbs free energies, ΔG.

Spin-spin coupling constant calculations at the B3LYP/aug-cc-pVDZ-su2[296] level (where the aug-cc-pVDZ-su2 basis set is a partially decontracted diffuse double-zeta basis set augmented with additional tight s functions) were performed on 3F-GABA snapshots from PM3/MM MD simulations using Gaussian 09. Snapshots were taken in intervals of 0.1 ps from various points of the unconstrained simulations (a total of 20-88 snapshots for conformers **D**, **J**, **C**; 250-430 for **E**, **H**, **G**, **I**, **A**; and more than 500 for **B** and **F**). Point charges from all surrounding water molecules were included in the NMR calculations.

4.2 Results: 3F-GABA MD simulations

4.2.1 Results: DFT/MM MD simulations

Initially, DFT/MM MD simulations were carried out at the B3LYP/6-31+G(d) and RI-B97-D[26]/6-31+G(d) levels. Simulations started from both folded and extended conformers (previously optimised at the PCM-B3LYP/6-31+G(d) level) and ran for ~20 ps. The results from these simulations were inconclusive, however, as the simulations only equilibrated around the starting minima. Ramachandran-type plots that show the conformational states of 3F-GABA on a 2D scatter plot as a function of the C^1-C^2-C^3-C^4 and C^2-C^3-C^4-N dihedral angles (see numbering in Figure 22), were used to analyse the results. A Ramachandran-type plot of the B3LYP simulations is shown in Figure 24.

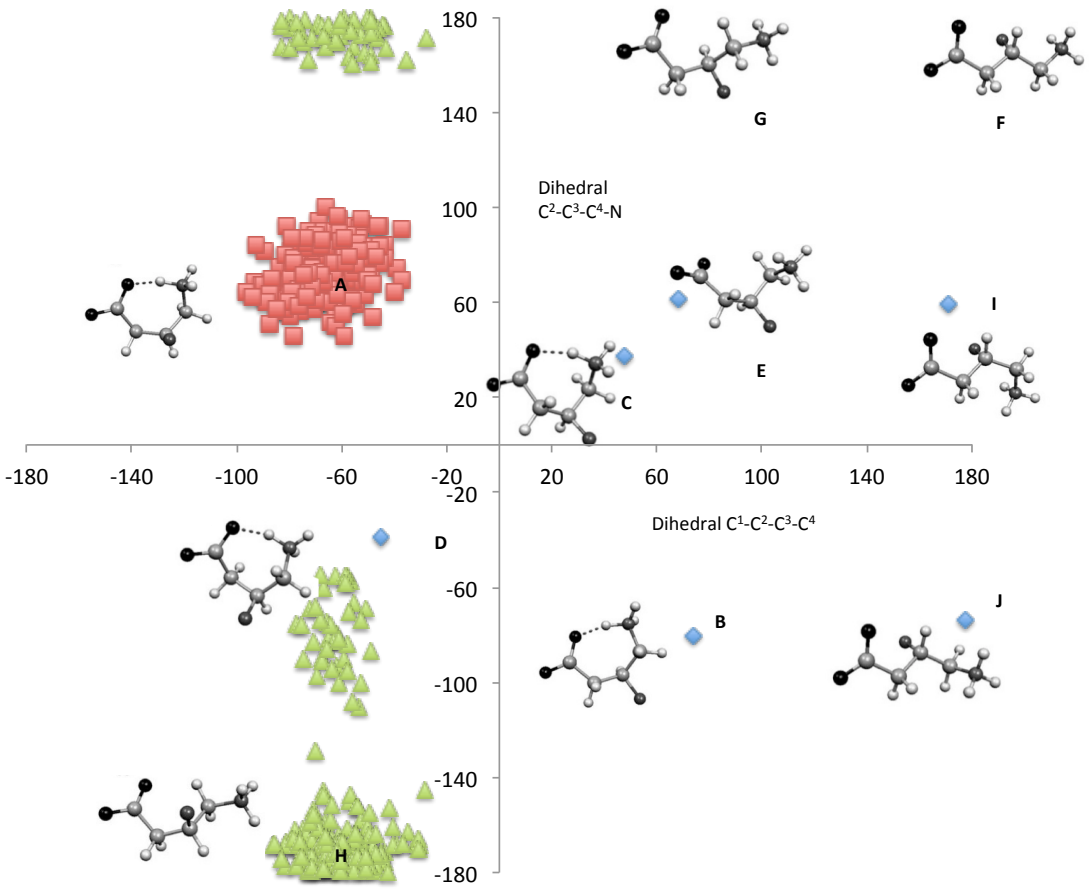


Figure 24 Ramachandran-type scatter plot showing the results of two B3LYP/6-31+G(d)/MM MD simulations. Squares and triangles indicate points visited after the two MD simulation starting from conformer **A** (squares) and distorted conformer **A** (triangles). Diamonds and labels indicate the position of PCM-DFT conformers on the plot.

Simulations using the much less expensive B97-D functional using the resolution of identity (RI) approximation[297-299] were performed in order to see if trajectories long enough to reach equilibrium were computationally tractable with DFT methods at all. The calculations employed the 6-31+G* basis set and the def2-SVP auxiliary basis set (present in the Turbomole basis set library), after making sure that the def2-SVP auxiliary basis set resulted in minimal artifacts. Constrained MD simulations were carried out (necessary for thermodynamic integration), constraining the C^2-C^3-C^4-N dihedral angle, and the force of constraint was sampled. A smaller active region of 442 water molecules was used as well to check if that would result in faster equilibration. Figure 25 shows the running average of the force of the constraint as a function of steps for one of these simulations.

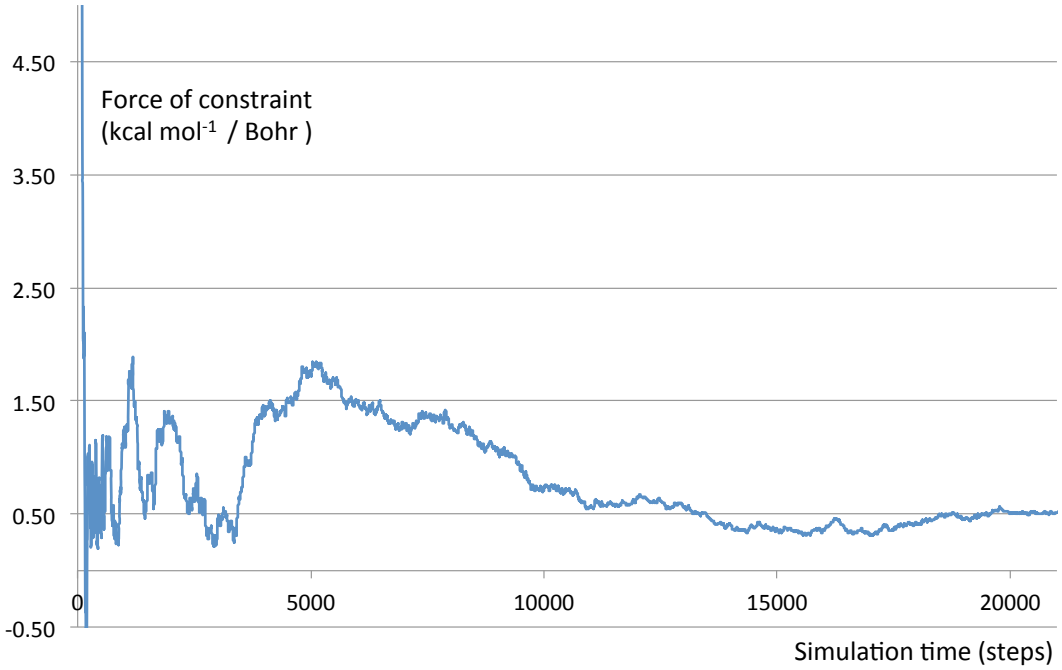


Figure 25 The running average of the force of constraint of dihedral C^2-C^3-C^4-N, as a function of simulation time (steps).

The results reveal that ~ 15 ps of simulation time (15000 steps) seem to be required before the force of constraint can be considered equilibrated.
This simulation required 18 days of simulation time on an 8-core node. As this is extremely expensive considering the number of simulations required for thermodynamic integration (several values of a reaction coordinate connecting two minima) we concluded that even RI-DFT would be too expensive for these simulations.

4.2.2 Results: Finding a cheaper QM method

As much longer simulation times were clearly needed, we opted for using a semi-empirical method as the QM method instead. A benchmark study was carried out, shown in Figure 26. Single-point calculations on PCM-B3LYP/6-31+G* optimised geometries were performed using different electronic structure methods. The high-accuracy WFT-based thermochemistry protocol G4MP2[300] (single-point calculations) was used to get very accurate relative energies between conformers to be used as reference values. Hybrid DFT methods such as B3LYP and M06-2X[24] were found to result in values close to reference values (MAE = 0.3 kcal/mol for B3LYP and MAE=0.4 kcal/mol for M06-2X) while GGA methods such as B97-D gave slightly worse results (MAE=0.6 kcal/mol). Semi-empirical methods performed considerably different with AM1[35] performing the worst (MAE=4.5 kcal/mol) while PM3[293] (MAE=1.5 kcal/mol) and PM6[301] (MAE=1.7 kcal/mol) gave much better results. As PM3 gave the best results of the semi-empirical methods and was available in the fast MNDO code it was considered the best option of these computationally cheaper methods.

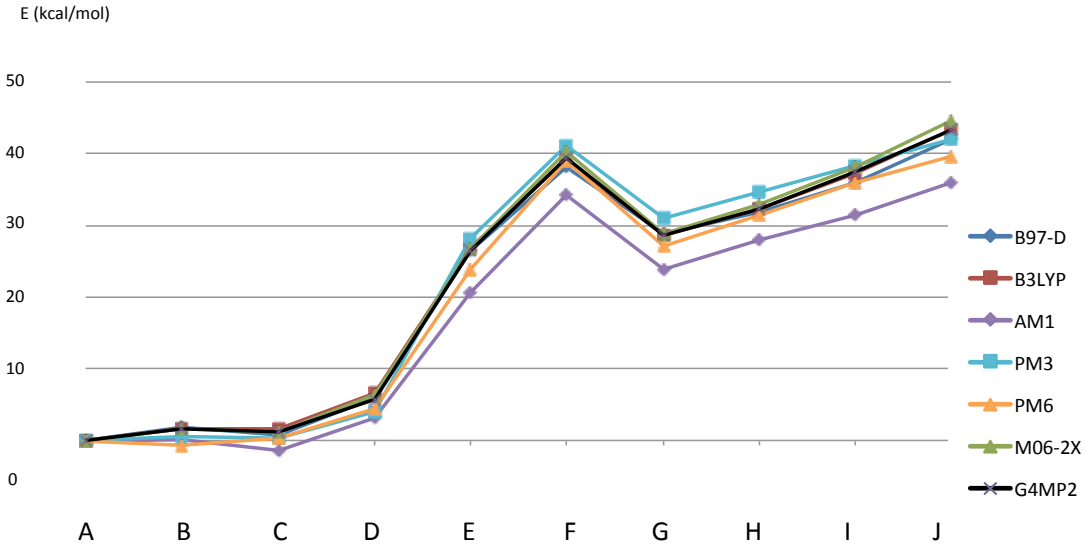


Figure 26 Relative gas-phase energies Δ*E* of 3F-GABA conformers with different QM methods. Single-point energies on PCM-B3LYP/6-31+G* optimised geometries with different methods.

PM3 optimisations with the PCM solvation model were carried out in order to see the effect of the continuum on the PM3 potential energy surface. The results are shown in Figure 27.

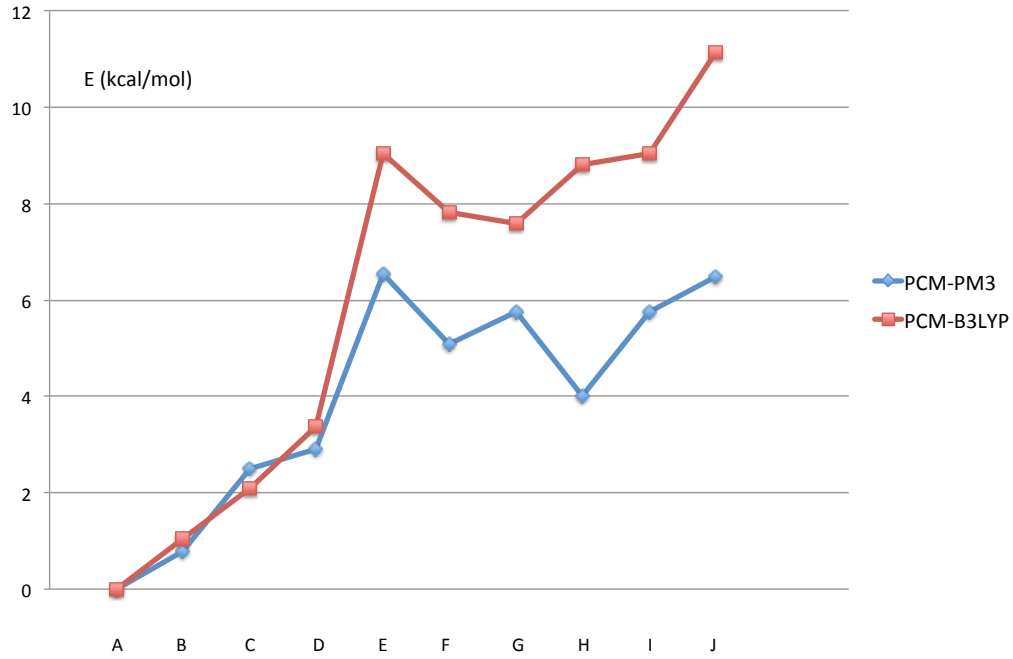


Figure 27 Relative energies Δ*E* of 3F-GABA conformers with PCM-PM3 and PCM-B3LYP/6-31+G* methods.

The PCM-PM3 results are slightly different than the the PCM-B3LYP results (MAE^{gas} =2.3 kcal/mol vs. MAE^{PCM}=1.7 kcal/mol , w.r.t. to B3LYP). However, qualitatively similar results are obtained: the lowest folded conformer A is more stable than extended conformers by 4-7 kcal/mol when including a PCM continuum in PM3 calculations.

4.2.3 Results: Unconstrained PM3/MM MD simulations

Several PM3/MM simulations were carried out with different starting conditions. Simulations were started after PM3/MM relaxation of the 3F-GABA minima that had been previously located at the PCM-B3LYP/6-31+G* level. The Ramachandran-type plots of the individual MD simulations reveal that the longer simulation times (up to 260 ps) allow the system to evolve through many conformers. Figure 28 shows the results of a 240 ps MD simulation, which started from conformer **J** (lower right of the plot). After visiting areas corresponding to conformers **G**, **F**, **I** etc., the folded form **A** was reached eventually. Unfortunately, even 260 ps of simulation time is not long enough for the system to reach equilibrium (in which case free energies could be assessed through relative populations of conformers in a Boltzmann distribution). Thus, only limited information can be deduced from these results. Of all conformers, **A**, **F**, **G**, and **I** are significantly populated in most simulations (100-260 ps long), thus suggesting that they may be low-energy conformers. In the simulation involving folded conformers, the intramolecular hydrogen bond seems to trap the system in a folded state for a considerable period. For simulations starting from conformers **B** and **C** the system ultimately escapes from a folded state and evolves towards extended conformers.

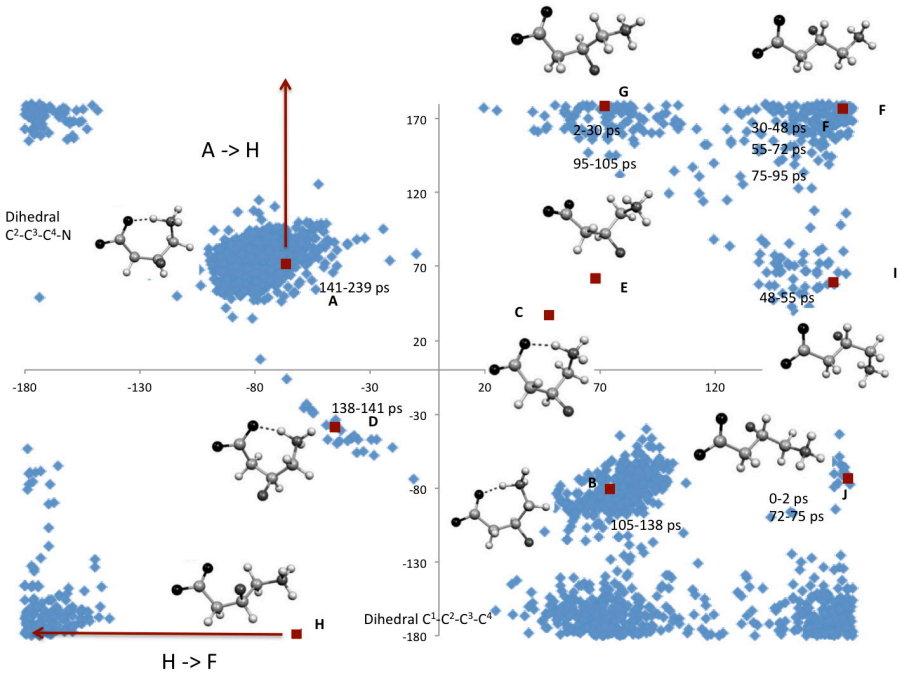


Figure 28 Ramachandran-type scatter plot illustrating the course of an unconstrained PM3/MM MD simulation of 3F-GABA in water. C^1-C^2-C^3-C^4 and C^2-C^3-C^4-N dihedral angles are taken as x and y axis, respectively. Diamonds indicate points visited after intervals of 100 fs. Squares with attached labels and 3D-plots denote the location of the PCM-optimised minima; approximate times during which an area around each minimum was visited are also included. Arrows denote reaction coordinates in subsequent constrained MD simulations.

In a few simulations, extended conformers also evolve into conformers **A**, **B** and **C** and back (in the case of **B** and **C**). In the simulation starting from conformer **A**, the simulation remains trapped in that folded state during the 260 ps simulation time.

This indicates a barrier to unfolding and that conformer **A** might be one of the lower-energy conformers on the PM3/MM energy surface.

4.2.4 Results: Constrained PM3/MM MD simulations

In order to shed light on the stability of folded vs. extended conformers in our QM/MM setup, free energy simulations were carried out using pointwise thermodynamic integration. The C^2-C^3-C^4-N dihedral angle was constrained and 18 simulations with dihedral values ranging from 75 to 177° were performed, starting from conformer **A**. This is equivalent to following a path from conformer **A** to the nearby extended conformer **H** (see the arrow in the top left quadrant of Figure 28). An analogous free-energy profile was also mapped between conformer **H** and **F**, this time constraining the C^1-C^2-C^3-C^4 dihedral angle in 15 simulations. The successive free-energy profiles for both **A**→**H** and **H**→**F** can be seen in Figure 29.

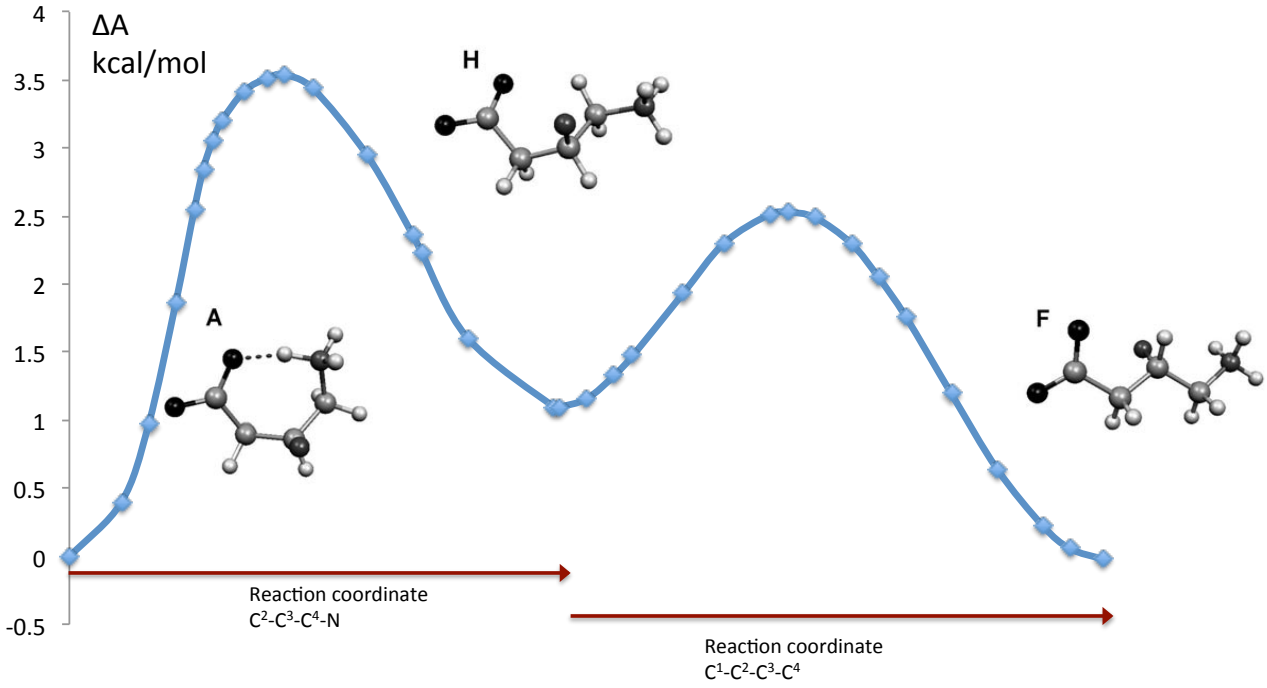


Figure 29 PM3/MM-simulated free-energy profiles for selected conformational changes (constrained MD for the reaction coordinates indicated by arrows in Figure 28).

Combining the resulting $\Delta A^{A->H}$=1.10±1.10 kcal/mol (**A** more stable) and $\Delta A^{H->F}$=-1.12±1.00 kcal/mol (**F** more stable) affords a free-energy difference between **A** and **F** of $\Delta A^{A->F}$= -0.02 kcal/mol. Our PM3/MM free-energy simulations hence predict the extended zwitterionic conformer **F** to be equally stable as the folded zwitterionic conformer **A**, in sharp contrast to PCM-PM3 calculations (see Figure 27) where **A** is 5.1 kcal/mol more stable than **F**.

As our QM/MM free-energy simulations were performed using the semi-empirical PM3 method, the question arises how much of an error the semi-empirical

approximation introduces to the calculation. Our gas-phase benchmark study (Figure 26) suggested that PM3 overestimates the energy of extended vs. folded conformers compared to DFT and high-level wavefunction theory (by ~1.8 kcal/mol for **F** and **A**). This information could be used to estimate a gas-phase correction to the PM3 solution data but instead we go one step further.

In an attempt to correct our PM3-obtained free-energy estimates to DFT-quality estimates we employ a thermodynamic cycle, shown in Figure 30.

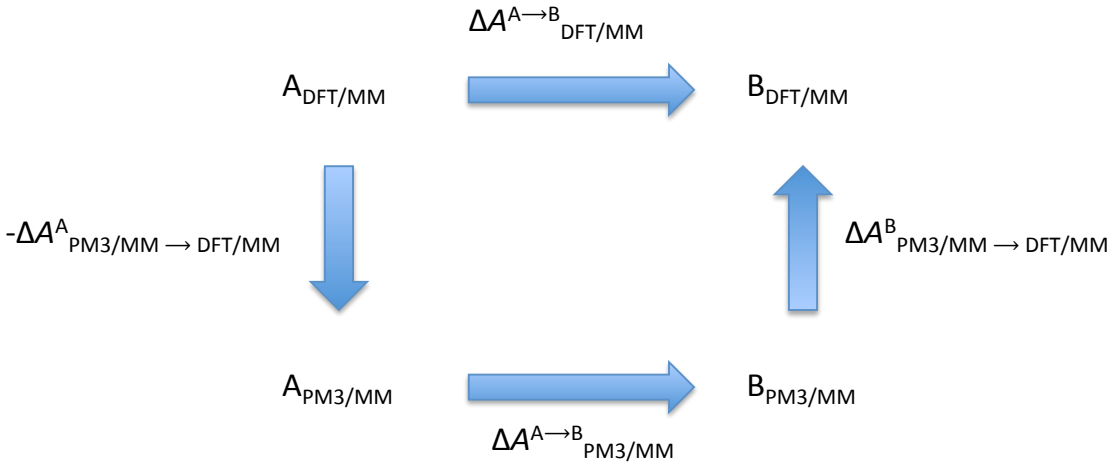


Figure 30 Thermodynamic cycle showing how the free energy difference between points A and B on a DFT/MM surface can also be estimated using PM3/MM energy differences and DFT corrections.

We thus calculate the $\Delta A^{A\rightarrow B}_{PM3/MM}$ term as accurately as possible using thermodynamic integration (as we have done) and then estimate the $\Delta A_{PM3/MM \rightarrow DFT/MM}$ terms by use of the single-step free-energy perturbation formula, Eq. 39, which now becomes:

$$\Delta A_{PM3/MM\rightarrow DFT/MM} = k_B T \; ln \; \langle e^{-(E_{DFT/MM}-E_{PM3/MM})/k_B T} \rangle_{PM3/MM(R)} \quad (68)$$

where single-point DFT energies are calculated on PM3/MM trajectories with PM3/MM energies subtracted.

This gives a a DFT level correction at low computational cost as long as the single-point DFT level energies are calculated at only a limited number of snapshots from PM3 level trajectories. The usefulness of the thermodynamic cycle and Eq. 40 rests on the assumption that the low-level QM surface and high-level QM surface are sufficiently similar as discussed by Warshel and coworkers.[302,303,304] Similar approaches of adding high-level corrections to low-level trajectories are beginning to become more popular in the literature, see for example, Refs.[305-307] and references therein.

The final DFT-corrected free energy difference is then:

$$\Delta A^{A\to F}_{DFT/MM} = \Delta A^{A\to F}_{PM3/MM} + (\Delta A^{F}_{PM3/MM\to DFT/MM} - \Delta A^{A}_{PM3/MM\to DFT/MM}) \quad (69)$$

1015 snapshots from PM3/MM trajectories for conformers **A** and **F** were used to calculate the ($\Delta A^{F}_{DFT/MM\to PM3/MM}$ - $\Delta A^{A}_{DFT/MM\to PM3/MM}$) term. The running average of $\Delta A_{DFT/MM\to PM3/MM}$ for A and F is shown in Figure 31 and already starts to show convergence after 600 snapshots.

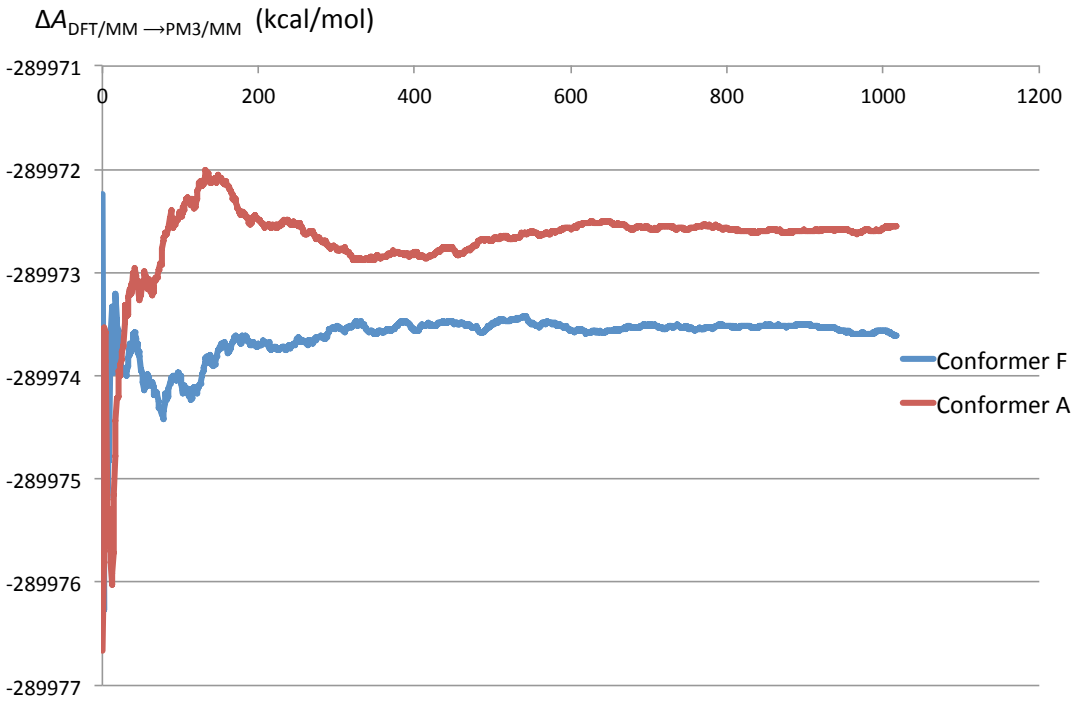


Figure 31 Running averages (in kcal/mol) of the PM3->DFT correction terms according to Eq. 68.

The ($\Delta A^{F}_{DFT/MM\to PM3/MM}$ - $\Delta A^{A}_{DFT/MM\to PM3/MM}$) term after 1015 snapshots is -1.05 kcal/mol, giving a final estimate for $\Delta A^{A\to F}_{DFT/MM}$ =-1.08 kcal/mol. Thus, the extended conformer **F** is indicated to be slightly more stable than folded conformer **A** by ca. 1 kcal/mol.

4.2.5 Results: Coupling constant analysis

The QM/MM free energy simulations suggest that extended conformers such as **F** are more stable than folded conformers like **A** in sharp contrast to previous continuum solvation model results.

We note, however, that the disagreement between PCM relative energies and experiment is based on a coupling constant analysis of an experimental NMR spectrum where PCM-DFT calculated coupling constants for folded conformers were found to be incompatible with the experimental coupling constant pattern while data for extended conformer **F** fitted better but with large deviations from experimental data nonetheless. The deviations were interpreted as two or more conformers contributing to the coupling constant pattern.

In view of this, we attempted a more elaborate coupling constant analysis using QM/MM data instead of PCM-DFT data. Our PM3/MM MD trajectories allow us to

calculate dynamically averaged coupling constants and using electrostatic embedding in the NMR calculation allows us to capture environmental effects on the coupling constants.

Using data from the PM3/MM trajectories, embedded single-point coupling constant calculations were carried out on a number of snapshots for each conformer, hence accounting both for explicit polarisation effects and dynamic effects on the coupling constants. This then allows one to deduce which conformers are contributing to the experimental spin-spin coupling pattern. We note that 3F-GABA is better suited to such a spectroscopic analysis than GABA as the former has a stereogenic centre, all backbone hydrogen atoms chemically and magnetically nonequivalent and has both $^3J_{HF}$ and $^3J_{HH}$ coupling data available. Table 29 shows the dynamically averaged $^3J_{HX}$ data from the PM3/MM simulations, obtained at the B3LYP/aug-cc-pVDZ-su2 level with embedded point charges included. The $^3J_{HH}$ couplings are a rather poor discriminator of the different conformers and will not be discussed.

In the previous study[281] calculations at the same level of theory were carried out but including the polarisable continuum model (as implemented in Gaussian 03) in the coupling constant calculations on PCM-B3LYP geometries. The original PCM-B3LYP data is given in Table 30 including RMS deviations from experimental data. Coupling constants were also calculated at the B3LYP//PCM-M06-2X level with and without including the polarisable continuum in the NMR calculations, now using the slightly different cavity definition in the PCM implementation in Gaussian 09; data are given in Tables 31 and 32. The results reveal the importance of including the polarisable continuum in the NMR calculations (RMS values decrease for the most part) but the different cavity definition and the different level of geometry optimisation have only small effects.

Table 29 Dynamically averaged ^3J(X,H) spin-spin coupling constants[a] from PM3/MM simulations.

	F-H^5	F-H^4	F-H^2	F-H^1	H^3-H^5	H^3-H^4	H^3-H^2	H^3-H^1	RMS,all	RMS,FH	RMS,HH
Exp[b]	14.6	34.0	17.0	31.3	9.5	2.3	8.1	4.0			
A	1.8(40)	24.8(68)	2.0(40)	36.3(73)	6.1(21)	0.8(10)	7.6(22)	1.1(11)	8.1	11.2	2.4
B	0.4(28)	11.1(58)	1.3(38)	17.1(73)	10.2(24)	1.8(18)	11.4(23)	1.7(15)	12.2	17.1	2.1
C	15.7(57)	32.4(59)	1.4(33)	19.2(83)	4.5(20)	1.2(11)	11.4(27)	1.4(3)	7.4	9.9	3.3
D	-0.1(12)	11.8(51)	21.9(120)	42.9(59)	10.2(28)	7.8(19)	5.9(27)	2.2(14)	10.6	14.7	3.1
E	2.8(45)	26.4(76)	2.0(39)	15.7(80)	5.7(24)	0.9(9)	12.7(31)	2.6(22)	9.4	12.9	3.1
F	10.1(64)	26.7(56)	14.8(91)	37.4(74)	10.2(27)	2.3(24)	11.3(28)	1.9(17)	4.1	5.4	2.0
G	13.0(71)	30.5(66)	4.5(66)	14.4(98)	9.7(28)	1.8(19)	11.7(29)	3.5(26)	7.7	10.7	1.8
H	10.7(64)	27.7(61)	6.3(66)	37.7(82)	10.6(32)	2.2(19)	6.7(29)	1.9(17)	5.2	7.3	1.4
I	3.0(45)	24.0(78)	14.9(88)	37.0(78)	5.9(25)	1.0(12)	12.5(3.0)	2.1(23)	6.2	8.3	3.1
J	3.4(62)	8.6(81)	8.7(73)	40.6(84)	9.9(26)	3.2(30)	12.0(21)	3.5(23)	10.9	15.2	2.0

[a]In Hz, B3LYP/ aug-cc-pVDZ-su2 level (in parentheses: standard deviations over all snapshots, in the units of the last digit - note that these indicate the sensitivity of each J-coupling to the dynamics, *not* statistical uncertainties).

[b]Experimental data and numbering scheme for H atoms from Ref.[281]

Table 30 Previous PCM-B3LYP coupling constant data (Hz) with RMS deviations.[281]

	F-H^5	F-H^4	F-H^2	F-H^1	H^3-H^5	H^3-H^4	H^3-H^2	H^3-H^1	RMS,ALL	RMS,FH	RMS,HH
Exp	14.6	34.0	17.0	31.3	9.5	2.3	8.1	4.0			
A	2.59	28.8	4.6	49.22	6.2	0.47	6.61	1.73	9.14	12.71	2.32
B	-1.4	13.97	4.43	15.53	10.46	2.04	11.22	2.73	11.60	16.31	1.76
C	17.27	33.99	1.23	22.58	4.45	1.02	10.48	7.91	6.89	9.11	3.47
D	-1.4	17.76	17.75	49.93	8.84	8.05	4.78	3.00	10.68	14.73	3.37
E	7.09	30.59	4.35	11.13	4.65	0.99	12.34	4.69	9.21	12.60	3.31
F	8.75	27.52	14.37	51.73	11.06	1.97	10.78	1.37	8.04	11.19	2.04
G	9.8	30.51	3.49	14.19	10.12	1.75	11.29	4.21	8.07	11.30	1.65
H	10.7	29.62	5.53	52.26	10.5	1.65	6.37	1.57	8.77	12.30	1.61
I	6.75	29.05	15.03	49.78	4.18	1.2	11.82	1.25	7.77	10.39	3.57
J	0.2	8.36	14.37	51.73	11.04	3.31	10.78	1.37	12.78	17.95	2.09

Table 31 B3LYP/aug-cc-pVDZ-su2//PCM-M06-2X/6-31+G* coupling constant data (Hz).

	F-H^5	F-H^4	F-H^2	F-H^1	H^3-H^5	H^3-H^4	H^3-H^2	H^3-H^1	RMS,ALL	RMS,FH	RMS,HH
Exp	14.6	34.0	17.0	31.3	9.5	2.3	8.1	4.0			
A	5.70	28.88	-0.81	36.64	4.33	1.05	9.31	0.28	7.87	10.62	3.30
B	2.36	5.30	-0.87	24.79	10.59	3.78	10.46	0.97	13.01	18.27	2.13
C	14.61	28.97	0.42	27.87	4.08	1.22	9.95	8.11	6.74	8.83	3.57
D	-1.49	16.62	18.15	47.90	8.31	8.08	4.55	2.94	10.53	14.47	3.48
E	5.21	23.13	5.77	10.13	4.91	0.74	11.67	5.22	10.11	13.97	3.07
F	5.84	15.43	19.76	62.07	10.91	2.67	9.79	1.36	13.17	18.55	1.73
G	6.53	23.07	5.06	15.34	9.39	2.22	10.26	4.69	8.56	12.06	1.13
H	8.18	19.63	5.33	62.71	9.31	1.99	6.75	1.10	13.14	18.51	1.61
I	4.35	17.84	21.11	52.11	4.09	1.06	11.48	1.17	10.41	14.28	3.54
J	1.86	2.50	12.69	49.60	10.80	4.43	11.20	2.41	13.81	19.42	2.14

Table 32 PCM-B3LYP/aug-cc-pVDZ-su2//PCM-M06-2X/6-31+G* coupling constants (Hz).

	F-H^5	F-H^4	F-H^2	F-H^1	H^3-H^5	H^3-H^4	H^3-H2	H^3-H^1	RMS,ALL	RMS,FH	RMS,HH
Exp	14.6	34.0	17.0	31.3	9.5	2.3	8.1	4.0			
A	7.07	35.52	-0.91	37.37	4.30	1.05	9.29	0.29	7.59	10.21	3.31
B	2.92	5.92	-0.97	24.58	10.61	3.81	10.64	1.00	12.81	17.98	2.18
C	16.74	33.94	0.45	24.80	4.01	1.26	10.06	8.16	6.83	8.95	3.62
D	-1.56	19.98	17.65	48.84	8.32	8.11	4.56	3.12	10.09	13.84	3.48
E	5.90	28.15	5.19	9.61	5.00	0.76	12.09	5.11	9.74	13.41	3.15
F	7.15	25.25	16.06	52.59	11.26	2.48	10.27	1.35	8.67	12.11	1.93
G	8.29	28.66	4.26	12.93	10.01	2.22	10.99	4.38	8.49	11.92	1.48
H	9.61	27.35	4.66	52.94	10.41	2.05	6.64	1.25	9.36	13.13	1.63
I	6.48	27.49	17.25	49.10	4.26	1.02	11.38	1.13	7.69	10.31	3.47
J	2.55	5.20	11.92	51.61	10.93	4.27	11.17	2.12	13.38	18.80	2.17

Table 33 M06-2X/MM optimised (only 3F-GABA allowed to move) coupling constant data (Hz) with 3F-GABA (only) in the QM region in the NMR calculation.

	F-H^5	F-H^4	F-H^2	F-H^1	H^3-H^5	H^3-H^4	H^3-H^2	H^3-H^1	RMS,ALL	RMS,FH	RMS,HH
Exp	14.6	34.0	17.0	31.3	9.5	2.3	8.1	4.0			
A	-0.13	19.80	9.14	46.02	8.25	0.13	5.11	2.89	9.44	13.20	2.03
B	0.53	23.53	1.62	14.36	7.51	0.09	11.12	2.34	10.32	14.41	2.27
C	14.33	34.65	-1.28	30.07	3.04	2.76	9.15	8.84	7.09	9.17	4.08
D	0.93	10.03	22.98	44.47	10.10	6.67	6.53	1.59	11.17	15.58	2.63
E	-0.70	10.85	2.64	16.35	9.31	0.64	12.49	3.12	12.36	17.32	2.39
F	9.97	32.25	16.07	41.72	10.84	2.28	2.92	2.68	4.53	5.79	2.75
G	15.16	33.40	4.33	12.33	9.87	1.10	10.46	6.53	8.17	11.41	1.84
H	14.53	35.14	5.75	50.58	10.84	1.28	6.65	1.68	7.98	11.17	1.61
I	7.03	31.41	8.40	44.49	5.17	0.67	12.73	2.67	6.68	8.83	3.34
J	0.46	9.71	8.79	49.67	10.63	3.33	10.83	2.70	12.28	17.28	1.69

Table 34 M06-2X/MM optimised (only 3F-GABA allowed to move) coupling constant data (Hz) using a larger QM region (water molecules included) in the NMR calculation.

	F-H^5	F-H^4	F-H^2	F-H^1	H^3-H^5	H^3-H^4	H^3-H^2	H^3-H^1	RMS,ALL	RMS,FH	RMS,HH
Exp	14.6	34.0	17.0	31.3	9.5	2.3	8.1	4.0			
A	0.32	24.13	8.73	45.19	8.66	0.53	5.59	3.29	8.46	11.86	1.63
B	1.10	25.49	1.66	14.27	8.04	0.45	11.52	2.70	9.99	13.96	2.17
C	15.69	38.61	-1.01	28.70	3.49	3.23	9.65	9.00	7.23	9.40	4.01
D	1.36	10.32	21.38	41.36	10.50	7.01	6.98	2.05	10.51	14.63	2.66
E	-0.43	12.54	2.35	14.79	9.87	0.97	13.03	3.61	12.25	17.13	2.57
F	10.71	36.81	16.13	38.03	11.86	2.78	3.46	3.30	3.48	4.16	2.64
G	16.51	37.76	4.17	11.09	10.49	1.54	11.02	6.81	8.72	12.15	2.12
H	14.62	41.71	4.84	45.73	11.63	1.78	7.10	2.26	7.28	10.19	1.49
I	7.91	36.12	7.72	45.25	5.82	1.10	13.22	3.09	6.82	9.08	3.24
J	1.13	11.77	8.69	51.25	11.31	3.92	11.57	3.08	12.05	16.90	2.17

Our dynamically averaged point-charge embedded data in Table 29 confirms that, based on RMS deviations of the $^3J_{HF}$ couplings, no folded conformer fits the experimental data and that the extended conformer **F** (RMS deviation of 5.38 Hz) shows better agreement with experiment than all other conformers. This is illustrated in Figure 32. The new PCM data on the other hand from Table 30 shows little discrimination between conformers, based on RMS values, as illustrated in Figure 33.

Additionally, coupling constant calculations with a larger QM region were performed. DFT/MM (M06-2X/6-31+G*/MM) geometry optimisations (only 3F-GABA allowed to move) were performed for each conformer (starting from a single snapshot of the PM3/MM simulations for each conformer). Subsequent coupling constant calculations were then performed using both a normal QM region and a larger QM region where

all water molecules within a radius of 4 Å of the 3F-GABA solute (96 in total) were included in addition to outer-shell point charges. The enlarged QM region would account for short-range quantum mechanical polarisation, not captured by point charges. The data is shown in Tables 33 and 34. Again, **F** is the only conformer to fit the experimental data in these calculations, and it is revealed that there is a small improvement in the RMS deviation in going from a small QM region (RMSD of 5.79 Hz) to a large QM region (RMSD of 4.16 Hz) for the $^{3}J_{HF}$ couplings of conformer **F**. These coupling constant data from a larger QM region is illustrated in Figure 34. It should be kept in mind, however, that the choice of a single snapshot for the DFT/MM optimisations may well introduce some bias.

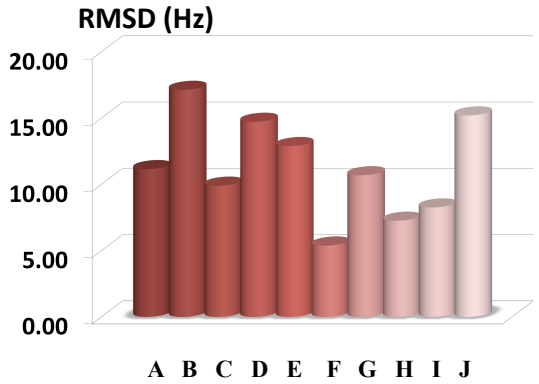


Figure 32 RMS deviations (Hz) in electrostatically embedded B3LYP calculated coupling constants on dynamically averaged data from PM3/MM trajectories.

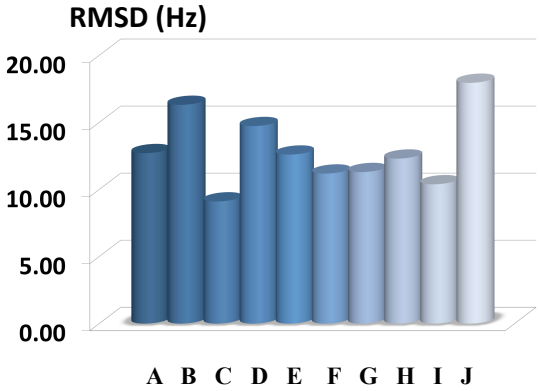


Figure 33 RMS deviations (Hz) in PCM-B3LYP calculated coupling constants for each conformer using PCM-M06-2X geometries.

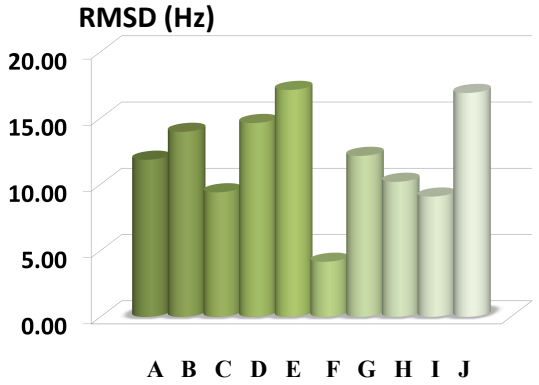


Figure 34 RMS deviations (Hz) in B3LYP calculated electrostatically embedded coupling constants using QM/MM optimised geometries using a large QM region.

The calculated NMR coupling constant data strongly suggest that conformer **F** is the dominant conformer in solution. If, as suggested by the free-energy simulation results discussed above, conformer **A** is higher in free energy than **F** by 1.1 kcal/mol, it should be populated by less than 14 % in an equilibrium mixture at 25 °C. The remaining discrepancy with experiment is presumably due to a combination of polarisation effects, basis set errors and the DFT functional error rather than being due to averaging of other conformers.

4.3 Results: Benchmarking continuum solvation models

All continuum solvation model calculations mentioned so far have used the original PCM variant, probably one of the most popular continuum solvation models in the literature. Other continuum solvation models and modifications to existing ones have been suggested over the years, for example: the conductor polarisable continuum model CPCM[308], isodensity PCM (IPCM),[79] self-consistent isodensity PCM (SCIPCM)[79] and the SMD solvation model.[78]

From our QM/MM MD simulations we evaluated a free energy difference between **A** and **F** at the B3LYP/6-31+G* level. The result is consistent with the NMR spectroscopic analysis but is in contrast to previous PCM-DFT results. We were curious whether other continuum solvation models were capable of predicting an energy difference between **A** and **F** closer to the QM/MM predictions.

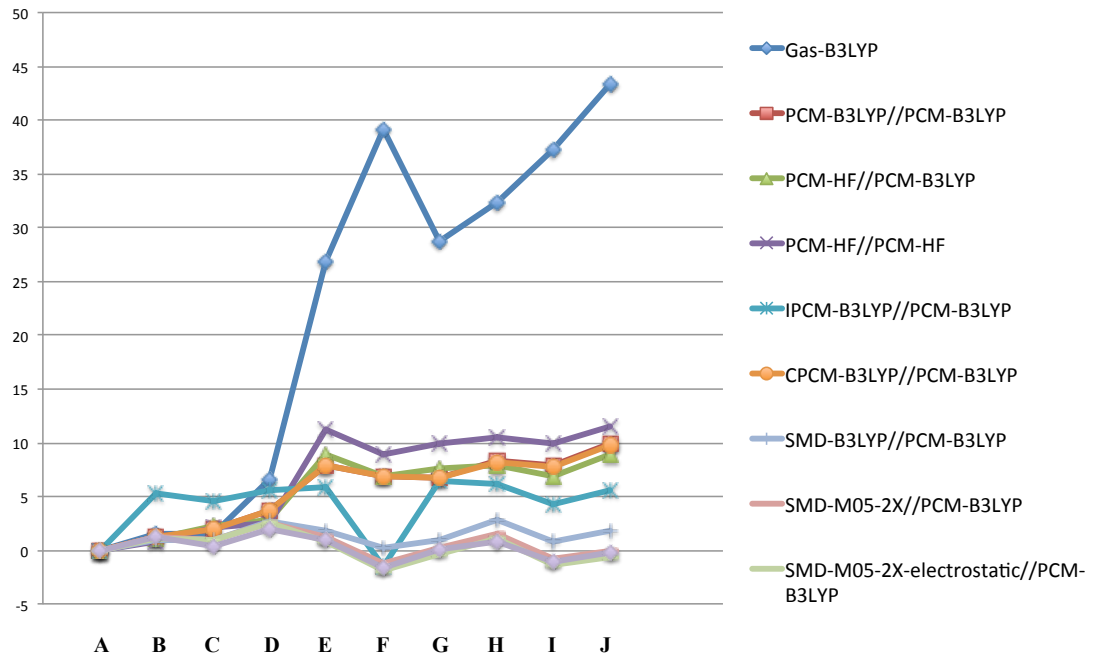


Figure 35 Relative energies $\Delta(E+G_{solv})$ of 3F-GABA conformers calculated by adding gas-phase relative energies (on PCM-B3LYP/6-31+G* geometries) and relative free energies of solvation from different continuum solvation models evaluated at the DFT or HF level with the 6-31+G* basis set.

The above mentioned solvation models, available in Gaussian 09, were used to calculate solvation free energies for each conformer, by subtracting total gas-phase energies from continuum solution energies, using the same level of theory for gas and solution. In order to do a consistent comparison between models, all solvation energies were added to gas-phase energies at the B3LYP/6-31+G*//PCM-B3LYP/6-

31+G* level (i.e. PCM-B3LYP geometries), giving $E+G_{solv}$ solution energies. SCIPCM results could not be included as most calculations did not converge (those that did converge, for **A**, **E** and **G**, showed very similar relative energies as regular PCM results).

$\Delta(E+G_{solv})$ relative energies for the 3F-GABA conformers, using several different solvation models are plotted in Figure 35. The figure also shows the relative gas-phase energies of each conformer where it is evident that the intramolecular interactions strongly favour folded conformers and that any preference towards extended conformers must come from an accurate calculation of the interaction between solute and solvent.

Results using the PCM solvation model at different levels of theory result in a dramatic drop in preference for folded conformers (compared to the gas-phase) but folded conformers are still favoured. Results using the PCM solvation model at different levels of theory and the CPCM solvation model are overall quite similar. The IPCM solvation model shows a slightly different energy surface and curiously the **F** conformer is considerably more stable than conformers **B**-**E** and **G**-**J** and is even slightly more stable than **A** in agreement with the QM/MM result. The more than 5 kcal/mol energy difference between **F** and other extended conformers, however, seems inconsistent with the results of the unconstrained MD simulations where most extended conformers were encountered regularly (although these were performed at the PM3/MM level of theory).
Results using the SMD solvation model show dramatically different results than other solvation models. There is a slight dependence on the level of theory used to calculate the solvation energy (B3LYP vs. M05-2X) but overall the results seem the most compatible with the results from the QM/MM MD simulations where conformer **F** is found to be slightly more stable than all other conformers (in agreement with the calculated energy difference and NMR analysis) but other conformers still being close enough in energy to be encountered during MD simulations. Also shown in Figure 35 are SMD results using only the electrostatic term (non-electrostatic term neglected) which shows that the performance of the SMD solvation model for our system has mainly to do with electrostatic effects but not the non-electrostatic (cavitation, dispersion, exchange-repulsion etc.) term which is of a slightly more empirical nature. The good performance of the SMD model compared to PCM (both methods use the IEFPCM protocol in Gaussian 09) thus must come from the atomic radii used in the SMD solvation model that have been specially optimised.[78]

Finally, we note that our $E+G_{solv}$ energies are not free energies of solution. This is due to the fact that we do not include ZPVE or thermal effects to enthalpy and entropy. These quantities should come from vibrational frequency calculations and it has been recommended to do these calculations in the gas-phase.[309] For 3F-GABA, however, the zwitterionic form is not stable in the gas-phase and vibrational frequency calculations thus cannot be performed. It is a current debate in the literature whether calculating vibrational frequencies using a continuum solvation model is a valid approximation to use for the evaluation of the free energy of solution.[309,310] For our system it is not clear whether the neglect of potentially crucial quantities like entropy is more dangerous than including an ill-defined thermochemical quantity. Test calculations of the thermal correction to enthalpy and entropy from vibrational frequencies in solution, however, lead to small differences between $\Delta(E+G_{solv})$ and $\Delta G_{soln} = \Delta(E+G_{solv}+G_{therm,sol.})$ relative energies.

We have shown that recent continuum solvation models are not in such strong disagreement with experiment (and QM/MM) for the 3F-GABA system. This bodes well for future studies of zwitterions as continuum solvation calculations are very straightforward to carry out and have a low computational cost compared to the orders of magnitude more involved QM/MM calculations. At least it is clear that continuum solvation models are very useful for an an initial exploration of the conformational energy surface of small molecules that interact strongly with the solvent.

4.4 Summary and Future Work

In our computational study of 3F-GABA we:

1. Provided additional computational spectroscopic data that strongly suggest extended conformer **F** to be the dominant 3F-GABA conformer in aqueous solution.
2. Used an elaborate QM/MM MD approach to calculate free energy differences between conformers of 3F-GABA that supports the interpretation of the experimental data, according to which the extended conformer **F** should be dominant in solution.

The microsolvated cluster calculations with up to 5 water molecules, with and without a surrounding continuum[293] were not successful in predicting a conformational energy surface with extended conformer **F** being lower in energy than **A**. It may thus be that even larger clusters are required to fully capture the solvent effects that influence the energy surface of a zwitterion such as 3F-GABA. As mentioned before, larger clusters make traditional geometry optimisations troublesome and a dynamic treatment may actually be needed for reliable results. Kamerlin et al. compared continuum models, mixed cluster-continuum models and QM/MM free energy simulations for the reaction of phosphate hydrolysis.[311] It was found that both continuum solvation calculation and full QM/MM free energy simulations arrived at the same result for the reaction barrier while mixed cluster-continuum calculations instead introduced artifacts. Accounting for the orientation and entropy of the water molecules was found to be important for proper behaviour of cluster models which becomes hard to do if one does not perform molecular dynamics.

Unfortunately, it is very difficult to separate free energies into enthalpy and entropy contributions (in principle possible by running many simulations at different temperatures and using the van't Hoff relation), and gauging the importance of electrostatic interactions with increasing number of water molecules would require many more simulations. It is thus not completely clear what physical effects are most responsible for stabilising the extended conformer **F**. Intramolecular hydrogen-bonding would stabilise folded conformers like **A**. Conformers that do not form an intramolecular hydrogen bond (**E**-**J**) would have both amino and carboxyl groups available for hydrogen bonding with the surrounding solvent, and if this results in an increased number of hydrogen bonds then this would presumably stabilise such conformers in solution relative to folded ones. We explored this possibility by calculating radial distribution functions for trajectories involving conformers **A** and **F**.

The calculated radial distribution functions are shown in Figure 36. Integrals over the N...O_{H2O} RDF up to r = 3.5 Å (an upper limit for hydrogen bonds) give a higher average of contacts for conformer **F** (5.1) than for **A** (3.9). For the O_{CO2}...O_{H2O} RDFs (one for each oxygen on the carboxylate group) the integrals give values of 4.2 and 4.3 for conformer **F**, while for conformer **A** we get values of 4.4 and 3.4. There is thus an increase by 1 in average N...O_{H2O} contacts and O_{CO2}...O_{H2O} contacts for the carboxylate and amino groups of 3F-GABA in going from conformer **A** to conformer **F**. This could be interpreted as 2 more solute-solvent hydrogen bonds that conformer **F** forms compared to **A** and after counting the intramolecular hydrogen bond that **A** forms but **F** does not, 1 hydrogen bond still remains for potential stabilisation of the **F** conformer. This could explain the stability of **F** over **A** although a more detailed analysis remains to be performed, ideally for each 3F-GABA conformer which could potentially be correlated with relative free energies.

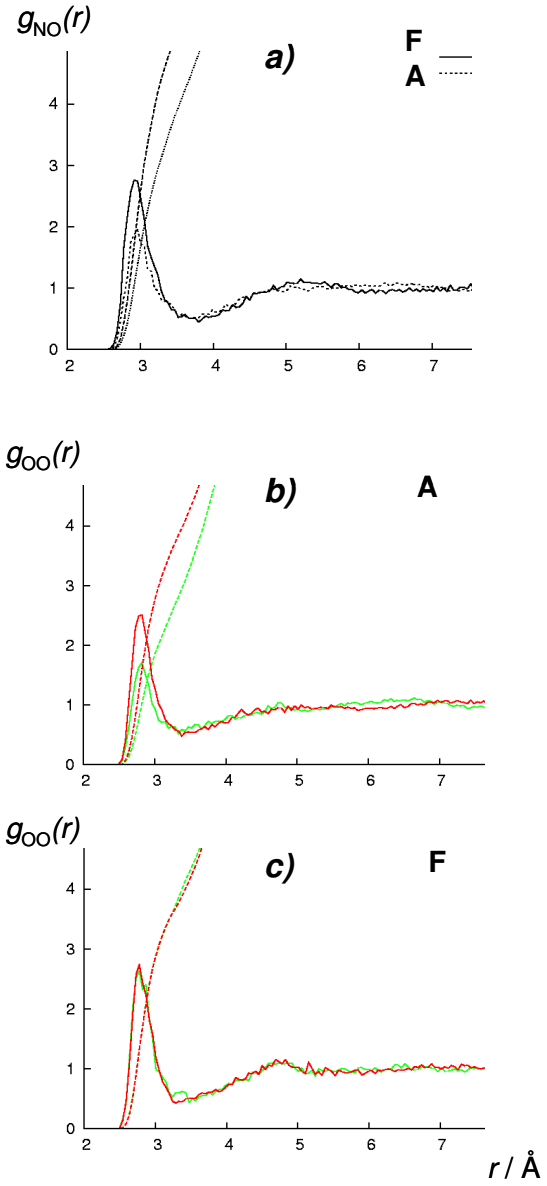


Figure 36 Radial distribution functions, g(r) between N of 3F-GABA and O atoms of the solvent (a) and between the carboxylate O atoms and the O atoms of the solvent (in b and c for each conformer). Dashed lines indicate integrals over g(r). From PM3/MM MD trajectories involving conformers **A** (2700 snapshots) and **F** (1300 snapshots).

In addition, extended conformers, like **F**, are less rigid and would be entropically more favourable than folded conformers, as discussed in a recent combined experimental and theoretical gas-phase study of neutral GABA conformers.[312] Also, since zwitterionic conformers become more stable with increasing number of water molecules and since the cationic and anionic parts are further apart in extended conformers this would likely result in more stabilising electrostatic interactions with the bulk solvent than for folded conformers in solution. Most likely, the conformational equilibrium of zwitterions such as 3F-GABA in solution, is a delicate balance of hydrogen bonding, solvent electrostatic effects and entropic effects.

We note that due to the complexity of the QM/MM MD simulations we have not fully explored the conformational energy surface of zwitterionic 3F-GABA. The standard deviations of the free energy differences obtained from the MD simulations are also quite high which one would like to reduce for greater confidence in the results. It is clear, however, that our QM/MM solvation model predicts the energy difference between conformers **A** and **F** to be completely different than that predicted by PCM solvation models as shown in Figure 37.

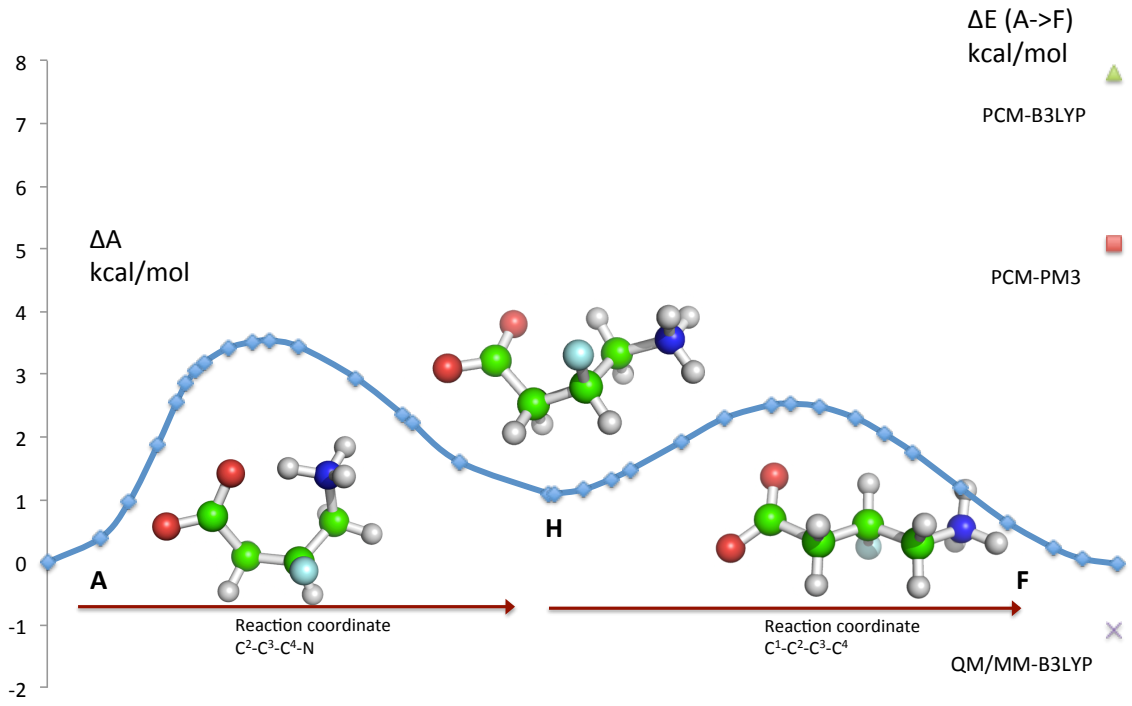


Figure 37 The computed PM3/MM free energy profile (line) with the PM3->DFT correction (QM/MM-B3LYP, marked as x) between **A**, **H** and **F**. Additionally, PCM computed potential energies of **F** (relative to **A**) with PM3 and B3LYP.

4.4.1 Possible improvements to QM/MM MD protocol

The QM/MM MD protocol for 3F-GABA that we ended up using, looks generally promising to model conformational changes of small molecules with non-negligible solvent effects, although being orders of magnitude more complicated than continuum solvation calculations. We now discuss some problems with the protocol and possible improvements that could prove useful for similar simulations in the future:

1. Contrained simulations wandering away from the important region:
In the constrained simulations we ran, we encountered problems when the simulations experienced a rare event jump from one part of coordinate space to another part of coordinate space that was unrelated to the free energy pathway being simulated. This is illustrated in Figure 38 where one can see how the constrained simulations with the C^2-C^3-C^4-N angle < 100° behave normally, equilibrating in the potential energy well around conformer **A**, with the C^1-C^2-C^3-C^4 angle varying between -120° and -30°. Other simulations behave similarly until a rare event (usually after a few tenths of ps) occurs, that makes them equilibrate on the other side of the Ramachandran-type plot (in the area of conformers **C**, **E**, **I**, **G** and **F**).

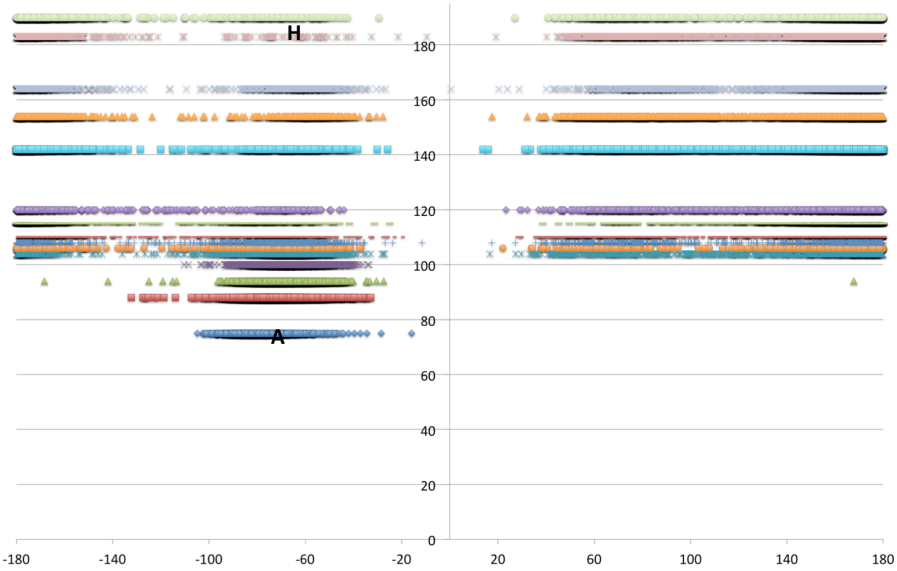


Figure 38 Constrained MD simulations of the path leading from **A** to **H**, represented as a Ramachandran-type plot. Y axis shows the constrained C^2-C^3-C^4-N angle in each simulation. X axis shows the variation in the C^1-C^2-C^3-C^4 angle.

This behaviour is unfortunate as we seek to model the shortest pathway between **A** and **H**, indicating that the simple dihedral angle is not an ideal reaction coordinate. In the limit of perfect equilibration throughout, this is not a problem, but any spontaneous processes that are irreversible on the timescale covered can introduce artifacts. One workaround is to select only those points that stayed on the direct path between **A** and **H**, but the necessary deletion of data before averaging complicates things considerably and introduces some bias.

In order to circumvent this nonideal behaviour, we explored adding a repulsive potential to the C^1-C^2-C^3-C^4 coordinate. As adding a repulsive potential means modifying the potential energy surface, the repulsive potential needs to be as well-behaved as possible i.e. not introducing artifacts. A harmonic potential with origin at x_0 =-74° was added to the simulations with force constant K. The value of the force constant (in Hartree/rad^2) was varied. Several MD simulations using a fixed constraint of 120 ° for the C^2-C^3-C^4-N and a restraint $V=\frac{1}{2} K (x-x_0)^2$ for C^1-C^2-C^3-C^4, were performed, varying the value of K. Figure 39 shows the variation of C^1-C^2-C^3-C^4 as a function of K. Potentials of K=30 and K=3 are clearly too narrow, restricting the dynamics of the molecule too much, while the potential with K < 0.0001 are too wide, allowing the molecule to equilibrate in the wrong part of the PES just as in the unconstrained simulations. It seems that potentials with K around 0.001 Hartree/rad^2 should be just right. The idea here is only to add some artificial

repulsion to the high-energy areas of ~0° and ~ -150° to prevent the molecule from crossing these barriers and escape to the right hand side of the Ramachandran plot, while not hindering the normal dynamics of the system. Adding a minimal restraint potential to this area seems to do the job although we note that in principle the direct effect on the mean force of the constraint should be looked at as well (only short simulations were performed here, hence we did not look into this).

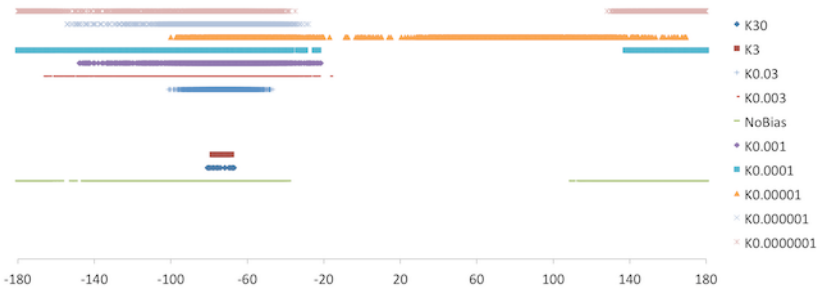


Figure 39 The effect of the force constant K (Hartree/rad^2) on the constrained C^2-C^3-C^4-N dihedral (y axis) MD simulations with added repulsive potential $V=½\ K\ (x-x_0)^2$ where x is the dihedral angle of C^1-C^2-C^3-C^4 (and is the x axis of the plot), x_0 corresponds to a dihedral angle of -74 °.

Future simulations of 3F-GABA/GABA or similar systems could thus explore using restraint potentials in this way, whenever this kind of behaviour is encountered.

2. The dihedral angle is not the best reaction coordinate

The choice of a dihedral angle as the reaction coordinate has an influence on how the constrained simulations behave and how quickly the simulations reach equilibrium. We explored other reaction coordinates as well that could connect the different 3F-GABA minima. Using atom distances as reaction coordinates is possible and using the distance between N and the carboxylic C was found to result in simulations with less fluctuation in the force (as seen through standard deviations). However, similar problems with simulations escaping to other regions of conformational space as in Figure 38 were encountered (although less severe). Still, clearly the choice of reaction coordinate deserves more attention. It is also possible that constraints like we use them are nonideal for this system and harmonic restraints should be used instead. This would then lead to the use of umbrella sampling[313] instead of thermodynamic integration to acquire free energy differences from results.

3. The approximate interaction between solute and solvent

In our QM/MM MD simulations the accuracy of the results will depend on how accurately the QM and MM regions interact with each other (compared to a full QM calculation of the full system). Since the QM-MM boundary does not cross a covalent bond the only interaction terms are the electrostatic term and the vdW term. The electrostatic term is calculated by elecrostatic embedding where the point charges from the MM force field polarise the electron density of the QM region. A physically motivated improvement over electrostatic embedding is polarised embedding where the point charges of the MM region change according to polarisation by the QM region, i.e. mutual polarisation by the two regions. Polarised embedding requires polarisable force fields that are beginning to become available and often also a self-consistent QM-MM polarisation cycle for each MD step. Van Gunsteren and coworkers recently explored QM/MM MD simulations using the polarisable GROMOS water force field describing solvent effects with promising results.[314,315]

It remains to be seen whether such polarisable embedded QM/MM MD calculations will be worth the effort as the self-consistent cycle required at each MD step will increase computational cost. Nevertheless, they may be a step towards more accurate QM/MM MD simulations of the liquid phase.

In traditional QM/MM calculations, the accuracy of the calculation can be made less dependent on the QM-MM interaction term, simply by increasing the QM region. In QM/MM MD simulations of a solvated molecule, increasing the QM region to include the nearest solvent molecules in the QM region, can create problems besides increasing computational cost. The reason for this is that the many interactions between solute and solvent molecules tend to be weak and shortlived in nature, meaning that water molecules in the QM region can be close to the solute at one instance and far from the solute a few picoseconds later. This would result in a fragmented QM region. QM/MM algorithms where QM and MM molecules can jump between regions have been developed [316-320]. In these adaptive QM/MM methods, a spherical region for example is defined around the solute and any solvent molecule that passes into the spherical region becomes a QM molecule and any QM molecule exiting becomes an MM molecule. In order for such simulations to behave smoothly, an extra region between QM and MM regions is often defined and a molecule in the "buffer" region will possess both QM and MM molecule characteristics. These methods are beginning to become more popular and have had increasing success but remain in a developmental stage.

We propose an alternative approach that could be explored for the 3F-GABA system or similar systems to test the dependence of the results of including QM water molcules around the solute. The MNDO code is a very fast semi-empirical code and in the QM/MM MD simulations, each (electrostatically embedded) QM energy+gradient evaluation takes less than 1 second on a single processor. The MM energy+gradient evaluation is in fact the bottleneck taking 17 seconds on a single processor. We explored increasing the QM region substantially to include all nearest neighbour water molecules and freezing the MM region entirely. The MM region is frozen in order to prevent exchange between QM water and MM water molecules as discussed before. We found that it was possible to increase the QM region up to 192 water molecules (MNDO has a maximum atom limit of 600 atoms), see Figure 40 for an illustration of such a cluster. In Table 35 we list calculation times for QM/MM gradient evaluations with differently defined QM regions and active regions.

Table 35 Calculation times (seconds) for QM/MM gradient evaluations with differing number of QM and active atoms.

No. H_2O in active region	No. H_2O in QM region	QM time	MM time	Total time
192	192	73	16.3	89.3
192	0	1	16.5	17.5
150	150	27	16.5	43.5
78	78	6.7	16.5	23.2
28	28	1.8	16.4	18.2
3061	0	1	16.5	17.5

Due to the not too steep scaling of the QM timings with system size, these large QM cluster calculations seem feasible. The smaller active region (equal to the QM region) will, however, create problems with the solute possibly equilibrating close to the active-frozen region boundary, leading to possible artifacts. We also note that the accuracy of PM3 to describe the many intermolecular interactions of such a cluster is questionable as hydrogen bonding and dispersion effects would have to be reliably described if the intent is to reliably improve upon the TIP3P force field and the QM-MM interaction term. However, we also note that according to a recent extensive benchmarking study, the OMX(X=1-3) semi-empirical methods[321] (available in MNDO) in combination with Grimme dispersion corrections are remarkably reliable and perform almost as well as GGA DFT methods for the whole test set of organic relative energies (even outperforming GGA and hybrid-GGA methods for some subsets involving weak interactions). The use of these semi-empirical methods in the fast MNDO code seems thus extremely promising and deserves further attention for use in QM/MM MD simulations, especially if thermodynamical cycles can reliably be used to move from semi-empirical to DFT free energy surfaces.

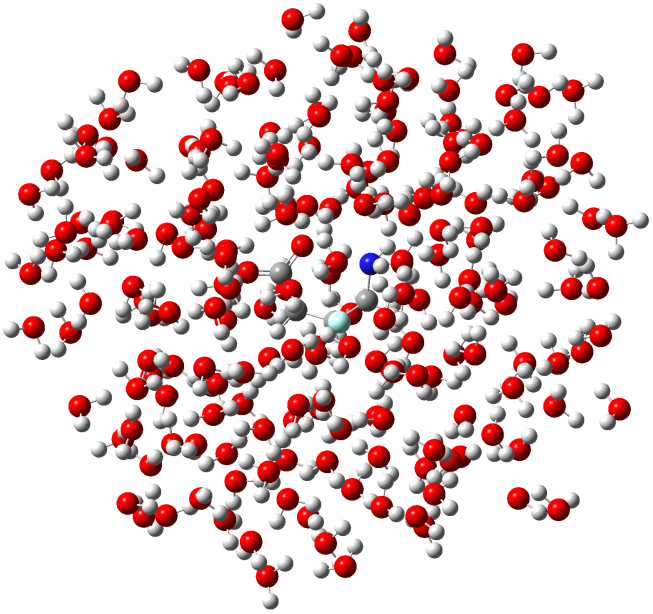


Figure 40 The 192 molecule hydrated 3F-GABA cluster.

4. Accuracy of the thermodynamic cycle

In our QM/MM free energy calculations we employed the thermodynamic cycle in Figure 30 and a single-step free energy perturbation step, Eq. 68, on a limited number of snapshots from PM3/MM simulations to estimate the $\Delta A_{\mathrm{PM3/MM} \rightarrow \mathrm{DFT/MM}}$ term. We do not know the reliability of this approximation for our calculations, it rests on the assumption that the DFT and PM3 potential energy surfaces are sufficiently similar in shape. A possible improvement of our protocol, however, would be to use the linear response approximation as defined by Warshel and coworkers[302] where $\Delta A_{\mathrm{PM3/MM} \rightarrow \mathrm{DFT/MM}}$ is approximated by calculating the PM3-DFT energy difference from snapshots of both PM3/MM and DFT/MM trajectories. This would increase the computational cost and complexity but could be useful to estimate what the accuracy of using a thermodynamic cycle really is.

5. Beyond constrained simulations

Finally, we mention that part of the complexity and problems of these QM/MM MD simulations stems from the fact that many different constrained simulations have to be run, along different reaction coordinates and then combined to yield free energy differences. In recent years, new algorithms to run MD simulations have been developed. One of these new algorithms is the metadynamics technique[322], designed to improve sampling of the important regions along a reaction pathway. A metadynamics MD simulation can be described as running a normal MD simulation while filling the free energy wells encountered with Gaussian bias potentials, thus slowly preventing the system from exploring regions of coordinate space that it has already explored sufficiently. A sufficiently long metadynamics simulation will have filled up all the wells on the surface defined by the collective variables with Gaussians and this information can in the end be added up to yield the free energy differences between minima. As the conformational energy surface of the 3F-GABA system seems to be well descibed by only two dihedral angles, it seems metadynamics could in principle be used to explore the full conformational free energy surface (FES) of

solvated 3F-GABA using the two dihedrals angles as collective variables. A thermodynamic cycle could also be used to go from a semi-emprical FES to a DFT FES. This remains to be explored.

Chapter 5. Isopeptide bond formation in a bacterial peptide: a QM/MM mechanism

Having studied structures, conformational equilibria and spectroscopic properties in the condensed phases we now explore mechanistic aspects of a reaction taking place inside a protein.

5.1 Introduction

Posttranslational modifications are chemical reactions that a polypeptide chain undergoes following the translation step in protein biosynthesis in cells.[323] This can include attachment of certain chemical groups, changes to individual amino acids (e.g. deimination, the process of converting arginine to citrulline), cutting of ends of the polypeptide chains and intramolecular reactions of amino acid sidechains. The most common of the latter type are disulfide bridges where thiol groups of cysteine residues of a protein react to form an S-S bridge by oxidation, usually promoted by oxidative agents. Recently, isopeptide bonds have emerged as yet another type of intramolecular reactions that a polypeptide can undergo.[324] An isopeptide bond is an amide bond, analogous to the peptide bond between amino acids in a peptide but it is formed between protein groups (sidechain or backbone) where one of the groups is not the α-amino or α-carboxy group.[324]
The best known isopeptide bonds are intermolecular, i.e. the isopeptide bond acting as a link between peptide chains. Examples are isopeptide bonds formed by ubiquitylation, sumoylation, transglutamination, sortase-mediated cell surface protein anchoring and pilus formation. All these examples are enzyme-catalyzed and involve the nucleophilic attack of the Lys ε-amino group to an α-carboxy group (except in transglutamination where Lys attacks the Gln carboxyamide group) proceeding through a thioester intermediate. An exception is the autocatalytic capsid assembly of the *E. coli* bacteriophage HK97 where Lys sidechains of one subunit attack Asn sidechains of other subunits turning into protein chainmail.[325]

The first *intramolecular* isopeptide bond was discovered in a one of the surface proteins (Spy0128) that form pili of the *Streptococcus pyogenes* bacterium in an X-ray crystallography study.[326] According to the electron density map, the bond was formed in two places, in both cases between the Lys and Asn residue sidechains in hydrophobic areas of the protein. Glu residues were in both cases found nearby, in hydrogen bonding distance to the isopeptide bond. Mutation studies were performed, substituting Glu with Ala, which resulted in no isopeptide bond formation, suggesting a catalytic role of the Glu residue.[327] Subsequent searches for the Lys-Asn-Glu triad in the Protein Data Bank has revealed that isopeptide bonds may be surprisingly common in many bacterial proteins and are present in all major pilus proteins, but have been missed in previous crystallographic analyses.[324]
A hydrophobic environment, the presence of Glu in hydrogen bonding distance and conserved positions in the protein fold seem to be common factors for the discovered isopeptide bonds so far.

A new type of isopeptide bond was recently discovered in the crystal structure of one of the domains of the FbaB surface protein of the bacterium *Streptococcus pyogenes* in the lab of Dr. Ulrich Schwarz-Linek at the University of St Andrews.[328,329] The FbaB protein can be divided up into several domains (most of which have not been characterised and may not be folded) and it is the folded CnaB2 domain where the

isopeptide bond was found by crystallography and will be discussed hereafter independently. The X-ray structure of CnaB2 revealed electron density between the sidechains of Lys and Asp residues, buried within the peptide. This is the first occurrence of the Asp residue instead of Asn as the partner to Lys in isopeptide bond formation.

Denaturation experiments were performed on the native protein and mutant versions which revealed that mutants lacking the isopeptide bond unfolded at relatively low temperature (53-59 °C) while the native form did not unfold appreciably until 100 °C. A mutant with the Asp residue turned into Asn (i.e. the same residue as in previous isopeptide bonds) resulted in isopeptide bond formation but at much slower rate (24 hours before completion of the reaction). The native protein was also found to be more stable with respect to lowering the pH.[329]

The presence of a nearby glutamic acid residue (as in previous cases of isopeptide bonds) and mutation experiments, suggested that the bond is formed autocatalytically in the protein, possibly directly catalyzed by the glutamate residue. The occurrence of this bond in the protein, its rapid kinetics (spontaneous formation) and relation to previous cases of isopeptide bonds spawned the following questions:

- How does this reaction occur ?
- What is the mechanistic reason for the lack of an isopeptide bond in the Glu->Gln mutation?
- Is there a common mechanism for the Lys-Asn and Lys-Asp reactions ?
- Why does the Asp->Asn mutation result in a slower isopeptide-forming reaction ?

In order to shed some light on these questions, we performed QM/MM calculations of the CnaB2 peptide. Part of this work was published in *Angewandte Chemie* in 2010 as part of a joint experimental-computational study on the structure, dynamics, thermal stability and isopeptide bond mechanism of the CnaB2 peptide.[329]

5.2 Computational details

5.2.1 Modelling an isopeptide bond

A prerequisite of most QM/MM studies on proteins is high-quality structural data. In this case, a high-resolution 1.65 Å crystal structure from X-ray experiments was available of the protein with the fully formed isopeptide bond.[328] As we wanted to explore the interaction of the residues in the hydrophobic pocket before reaction, the isopeptide bond was cleaved. This meant moving the residues a bit apart and adding the missing oxygen atom (due to water dissocation) on Asp. This also meant deciding on the number of protons present before the reaction and assigning an initial protonation state. As the crystal structure shows an O-O distance of 2.96 Å between the O atom of the amide bond and one of the O atoms of Glu this clearly indicated a hydrogen bond and hence a missing hydrogen atom (see Figure 41).

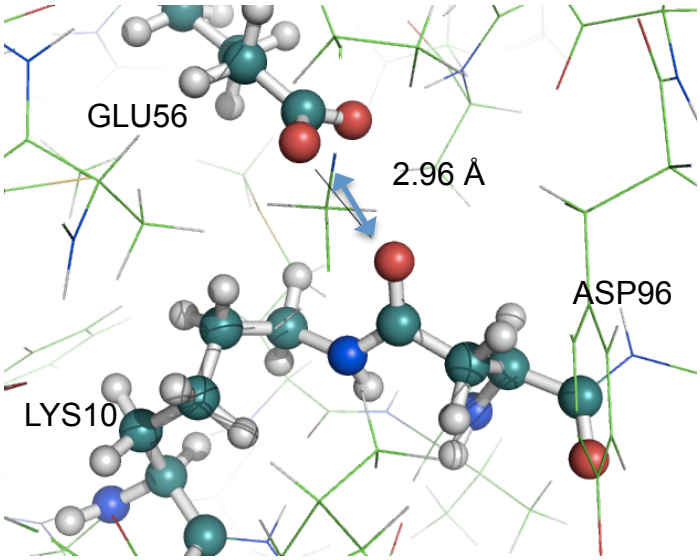


Figure 41 The hydrophobic pocket of the crystal structure of CnaB2 showing the isopeptide bond between Lys10 and Asp96 and the hydrogen bond to Glu56.

As two hydrogen atoms dissociate as part of a water molecule and one hydrogen remains on the N atom of the amide, this added up to a a total of 4 "active" hydrogen atoms among the residues after reaction. We assumed 4 hydrogens would exist in the pocket before the reaction as well. The lysine was assumed to be initially protonated (NH_3^+) and forming a saltbridge to a nearby carboxylic acid ($RCOO^-$), here the aspartate. This meant the glutamate had to be protonated (RCOOH) to account for all 4 hydrogens (see Figure 42). Other combinations were also explored but this combination was eventually settled upon as an initial state of the system. Other protonation states were subsequently explored as part of the QM/MM calculations.

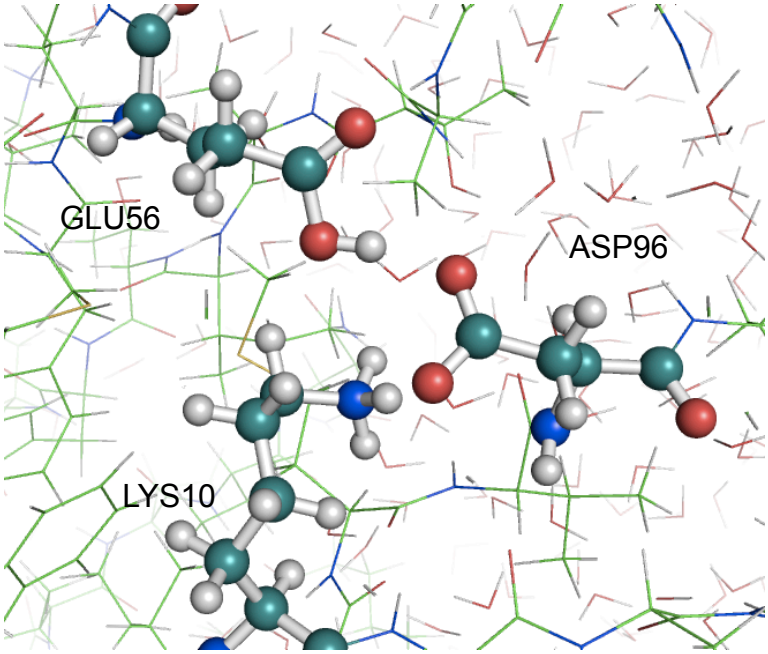


Figure 42 Modified crystal structure of CnaB2 with the isopeptide bond cleaved and missing oxygens and hydrogens added to the residues.

5.2.2 Protein preparation

QM/MM studies of proteins typically require a comprehensive preparation that is usually performed at the molecular mechanics level.[80] The following preparation protocols were used.

1. Main preparation

The X-ray coordinates of the protein (PDB code 2X5P) were imported into the CHARMM program[273] (version 34b2) and the protein was prepared using the CHARMM22 protein force field.[81,82] The sequence was renumbered, so that the N-terminus residue Asp22 (X-ray numbering) became Asp1 (our numbering) etc. The amide bond was cleaved into protonated Lys10, protonated Glu56 and unprotonated Asp96. Protonation states of other titratable residues were assigned based on calculated pK_a values from the Propka program[330] and chemical intuition (aspartate residues 1, 14 and 89). The two histidines were given CHARMM protonation states HSE (defined as the farther imidazole nitrogen being protonated). Missing hydrogen atoms were added and neutrality of the protein was imposed by selective protonation of surface aspartate residues. The protein was solvated (see Figure 43) by placing it in a water sphere with radius 35 Å (water molecules described by the TIP3P force field[274]), deleting clashing water residues (those whose atoms came within 2.8 Å distance to a protein atom). Short classical optimisations and MD simulations were run and more water spheres were added in the same place and a loop of deletion and equilibration went on for a few cycles until only relatively few water molecules could be added in each cycle. The total system size was 18507 atoms after solvation. The classical protein preparation finished by a 500 ps classical MD simulation during which all atoms were active except for an outer layer of solvent (defined as being further than 5 Å from a protein atom). Snapshots from the 500 ps trajectory were extracted and used for QM/MM calculations.

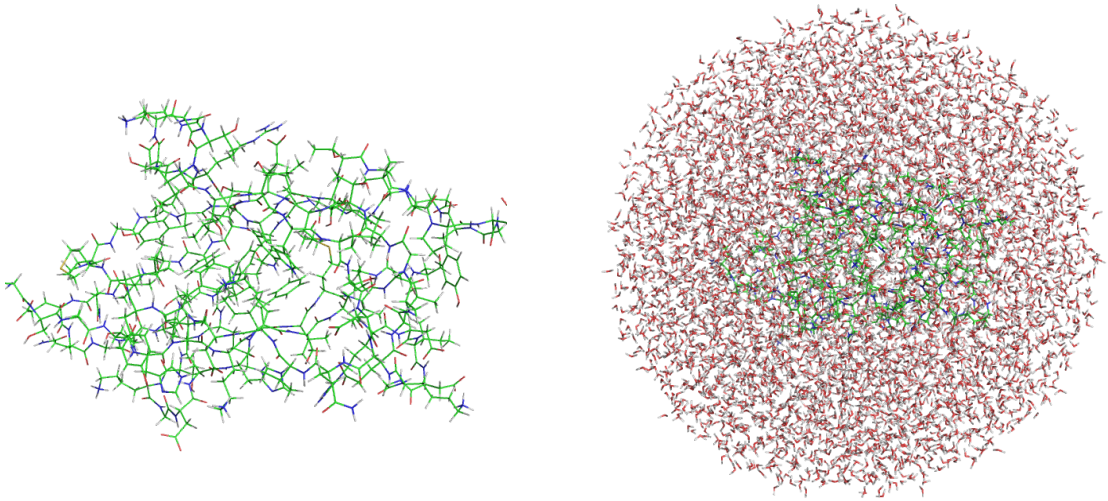


Figure 43 Molecular structures of all-atom CnaB2 in vacuum and fully solvated.

2. Intact isopeptide bond preparation

The isopeptide bond was kept intact and a topology for the isopeptide bond was defined analogous to a normal peptide bond in CHARMM. The protein was then prepared in the same way as the main preparation (above) but the final equilibration step was run for 1 ns. This preparation was done solely to check the reliability of the

force field to describe the main structure of the protein. A running average of the structural root mean square deviation (w.r.t. to the heavy atoms of the X-ray structure) of snapshots every 1 ps from the MD simulation of the protein, very quickly (after 50 ps) stabilises at ~1.0 Å and remains at ~1.0 Å during the 1 ns simulation. This suggests that the CHARMM force field describes the protein structure quite well.

3. Frozen QM/MM minimum preparation

Starting from a QM/MM optimised structure along the pathway (see later) that was prepared using the main preparation, the protein was reimported into CHARMM, most of the Lys10, Glu56 and Asp96 atoms were kept frozen and a 1 ns MD simulation was run.

4. Glu->Gln mutant preparation

The protein was prepared in the same way as in the main preparation except Glu56 was mutated into Gln56; Lys10 was assumed to be protonated and Asp96 deprotonated, forming a Lys-Asp salt-bridge.

5.2.3 QM/MM calculations

Snapshots from the classical protein preparation were imported into Chemshell where all QM/MM calculations were performed. Two QM regions were defined: one consisting of 30 atoms and the other of 62 atoms. QM-region I is a minimal region that only includes the atoms necessary for the reaction to take place. Trial calculations with this QM region, however, indicated that it was to small as some protonation states of residues along the reaction pathway only became stable with the larger QM-region II. The two QM regions shown in Figure 44 include link atoms.

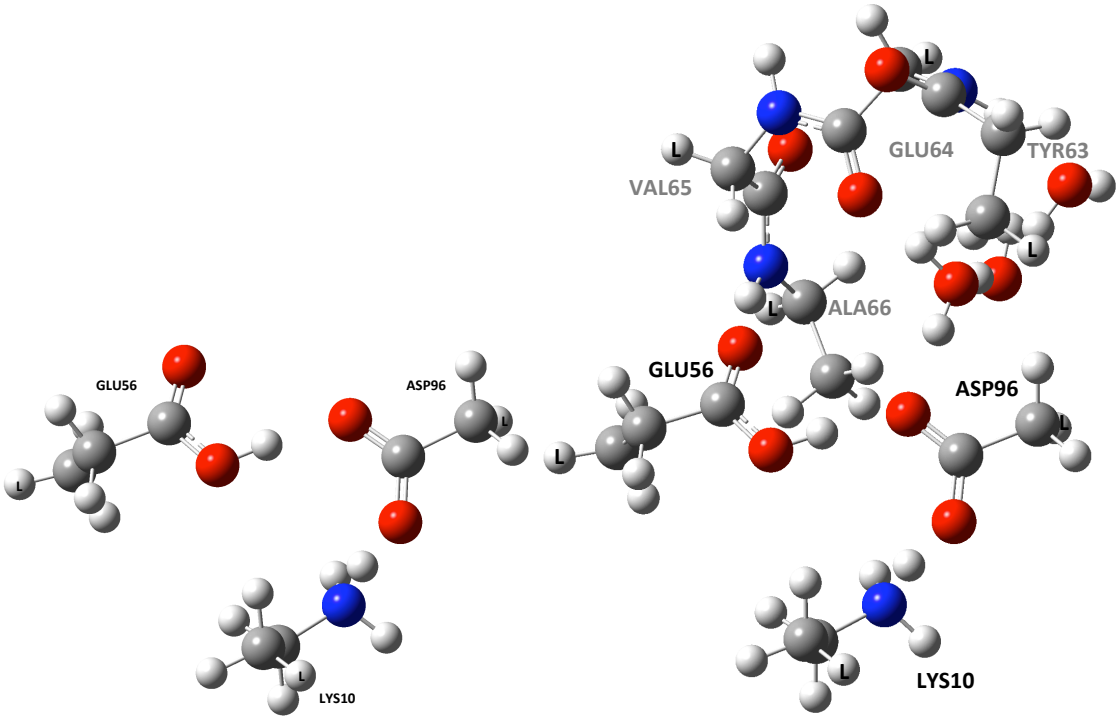


Figure 44 Left: QM-region I (30 atoms) Right: QM-region II (62 atoms). Link atoms indicated.

An active QM/MM region was defined consisting of all atoms (930) within an 12 Å radius of the Lys10 amino group. All other atoms were kept frozen.

Turbomole[276] version 5.10 was used as the QM code using both hybrid B3LYP and GGA B97-D functionals (RI approximation[297-299] used for B97-D calculations). The 6-31+G(d,p) basis set was used for B3LYP and RI-B97-D calculations with the auxiliary basis set aug-cc-pVDZ (from the Turbomole library) used as auxiliary basis set for the RI-B97-D calculations (the accuracy of the auxiliary basis set was tested and found to be satisfactory). DL_POLY[277] was used as the MM code (using imported CHARMM force field parameters from the classical protein preparation). An electrostatic embedding scheme was used in combination with link atoms and the charge-shift scheme as implemented in Chemshell.[88] Geometry optimisations were performed both using HDLCOpt[275] and the DL-FIND[331] geometry optimisation programs in Chemshell. Hybrid delocalised internal coordinates (HDLC)[275] were used as the coordinate system. Transition states were located by a combination of constrained optimisations along a bond coordinate with subsequent eigenvalue-following TS searches as well as combinations of nudged elastic band path optimisations[332-334] and subsequent TS searches with the dimer method.[335-338]

Single-point calculations of minima and transition states were calculated at higher DFT levels using the B3LYP and M06-2X functionals and the 6-311+G(3df,3pd) basis set by separate embedded single-point calculations using Gaussian 09.

5.3 Results: Initial QM/MM mechanism of native CnaB2

Using QM region II (62 atoms) we explored the QM/MM potential energy surface leading to isopeptide bond formation for the 500 ps snapshot of the main protein preparation. Atoms of the QM region were manually moved around in order to locate new minima and transition states to find the lowest energy pathway. The lowest energy pathway located is illustrated in Figure 45.

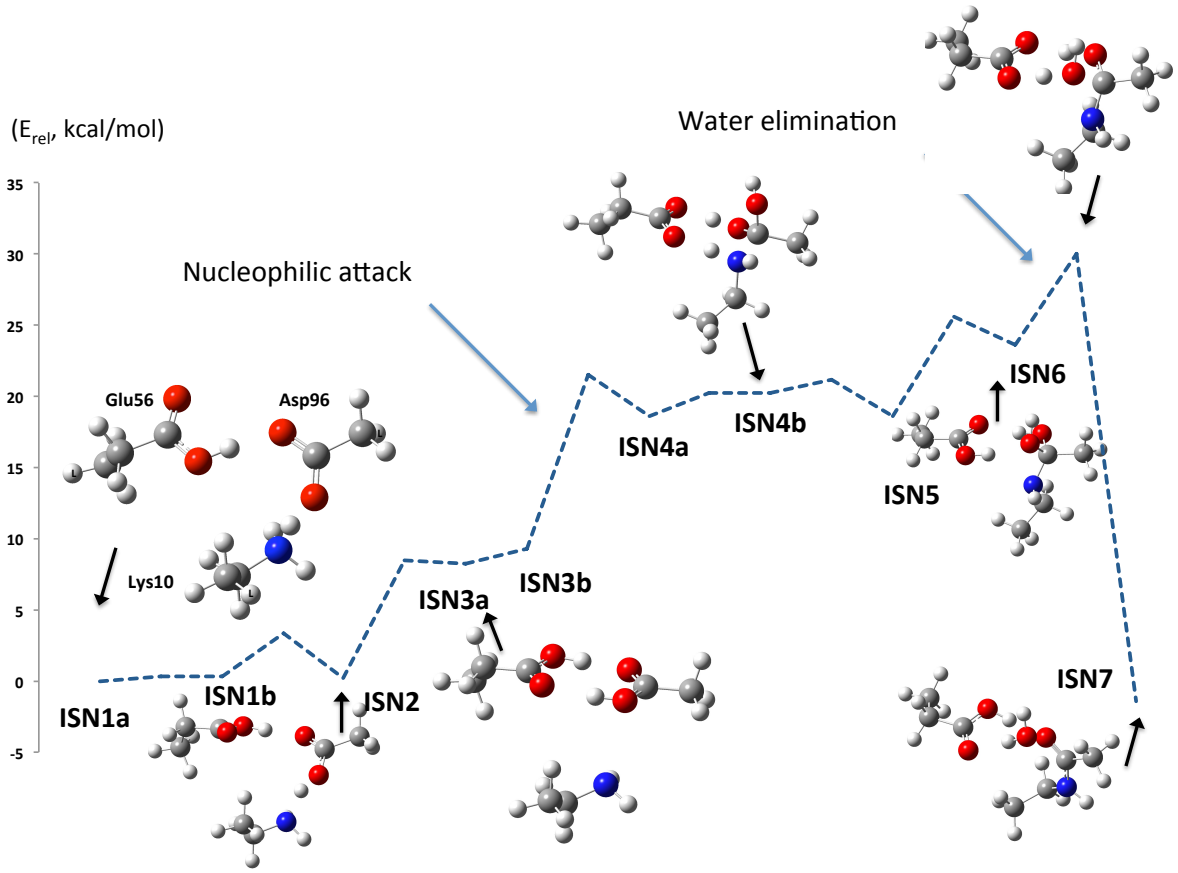


Figure 45 The isopeptide reaction pathway illustrated according to calculated minima and transition states. Only part of the QM region used in the calculations is shown. Calculated at the B3LYP/6-31+G(d,p)/MM level.

The pathway is described as follows: Starting from the structure from the initial QM/MM optimisation, **ISN1a**, a proton transfer from Lys10 to Asp96 occurs essentially without energy penalty leading to **ISN1b**, the Glu56 carboxylic acid group must then flip for correct orientation (the classical MD simulation showed this group flipping repeatedly) leading to **ISN2**. The still intact hydrogen bond between Asp96 and Lys10 is next broken and redirected towards Glu56, **ISN3**, leaving Lys10 essentially without hydrogen-bond partners. The energetic penalty for this is lower than expected, presumably due to the formation of the favourable carboxylic acid dimer orientation between Glu56 and Asp96. After a slight rotation of the NH_2 group, **ISN3b** deprotonated Lys10 is now set up in a favourable position for nucleophilic attack to Asp96.

The nucleophilic attack (see Figure 46) proceeds through a barrier which is presumably partly stabilised by a Lys10-Glu56 hydrogen-bond that is formed at the same time as the Asp96-Glu56 hydrogen-bond breaks and the N-C bond is formed. The resulting tetrahedral intermediate **ISN4a** is zwitterionic as the nitrogen has four bonds and one of the carboxylic oxygens of Asp96 is unprotonated (but forms a strong hydrogen bond with Glu56). Protonating the Asp96 carboxylic oxygen by Glu56 is a slightly uphill step and moves the anionic part of the zwitterionic state from Asp96 to Glu56 and creates an amide hydrate intermediate, **ISN4b**. A proton transfer from the NH_2 group of Lys10 to Glu56 results in a neutral amide hydrate intermediate **ISN5** that is slightly more stable than **ISN4a** and **ISN4b**.

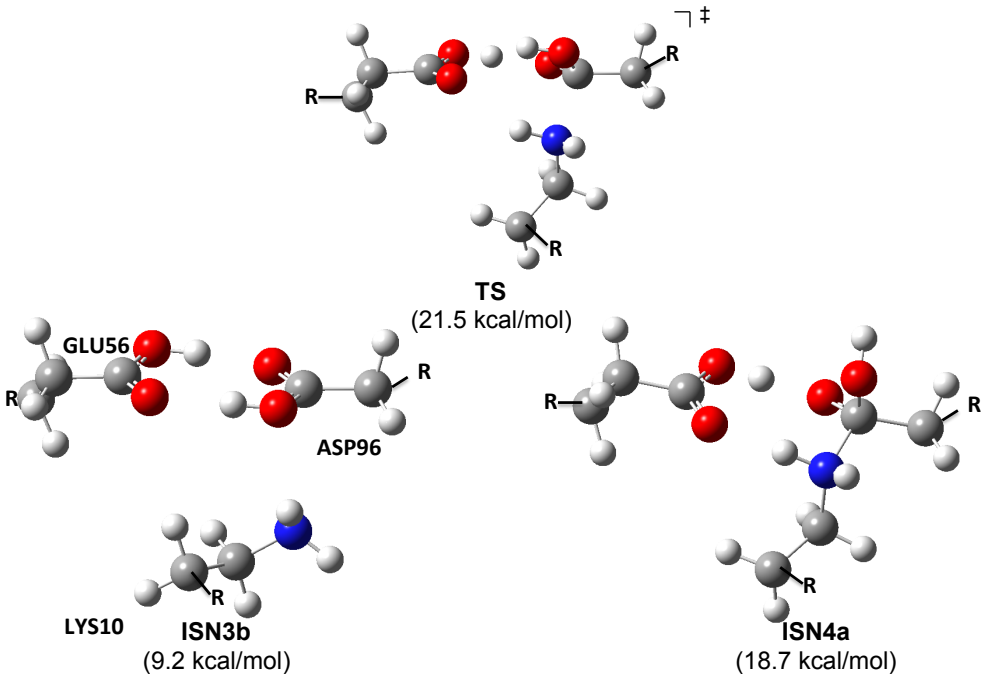


Figure 46 The nucleophilic step as predicted by QM/MM calculations. Only part of the QM region used in the calculations is shown. Relative energies of minima and the transition state w.r.t. to the reactant state **ISN1a**. B3LYP/6-31+G(d,p)/MM calculated. R indicates position of the link atom and the connection to the MM region.

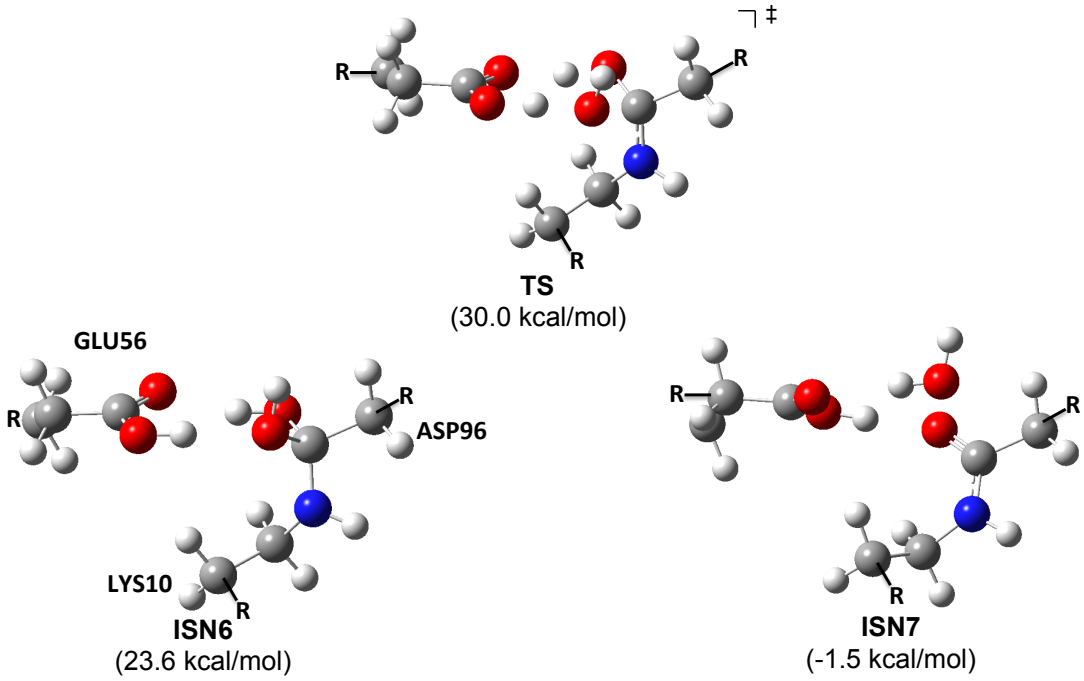


Figure 47 The water dissociation step as predicted by QM/MM calculations. Only part of the QM region used in the calculations is shown. Relative energies of minima and transition state w.r.t. to the reactant state **ISN1a**. B3LYP/6-31+G(d,p)/MM calculated. R indicates position of the link atom and the connection to the MM region.

The last step of the amide reaction is water dissociation (see Figure 47) of the now combined Lys10-Asp96 residue. The most direct way to achieve this is by redirecting the Glu56 hydroxyl group (hydrogen bonding with Lys10) towards the hydroxyl group of the Lys10-Asp96 residue (intermediate **ISN6**). The water dissociation can then proceed through a single step where the hydroxyl group is protonated by Glu56 and Glu56 is reprotonated by the other hydroxyl group of the Lys-Asp residue to give the isopeptide bond residue, dissociated water and protonated Glu56.

The reaction profile shown in Figure 45 was calculated at the B3LYP/6-31+G(d,p)/MM level and is rather high in energy. The water dissociation step is the highest point on the profile and has a relative energy of 30.0 kcal/mol. According to transition state theory this barrier is too high to be surmountable by the system at room temperature meaning that our calculations are at odds with experimental observations. As other mechanisms had been trialled and were found to be even higher in energy than the current mechanism, we explored other reasons for these high energy differences.

We performed single-point calculations with a much larger basis set, 6-311+G(3df,3pd) which should be close to the basis set limit. This, however, resulted in only small changes (see Figure 48) and even resulted in higher energies for some intermediates and a still high activation barrier of 30.7 kcal/mol.

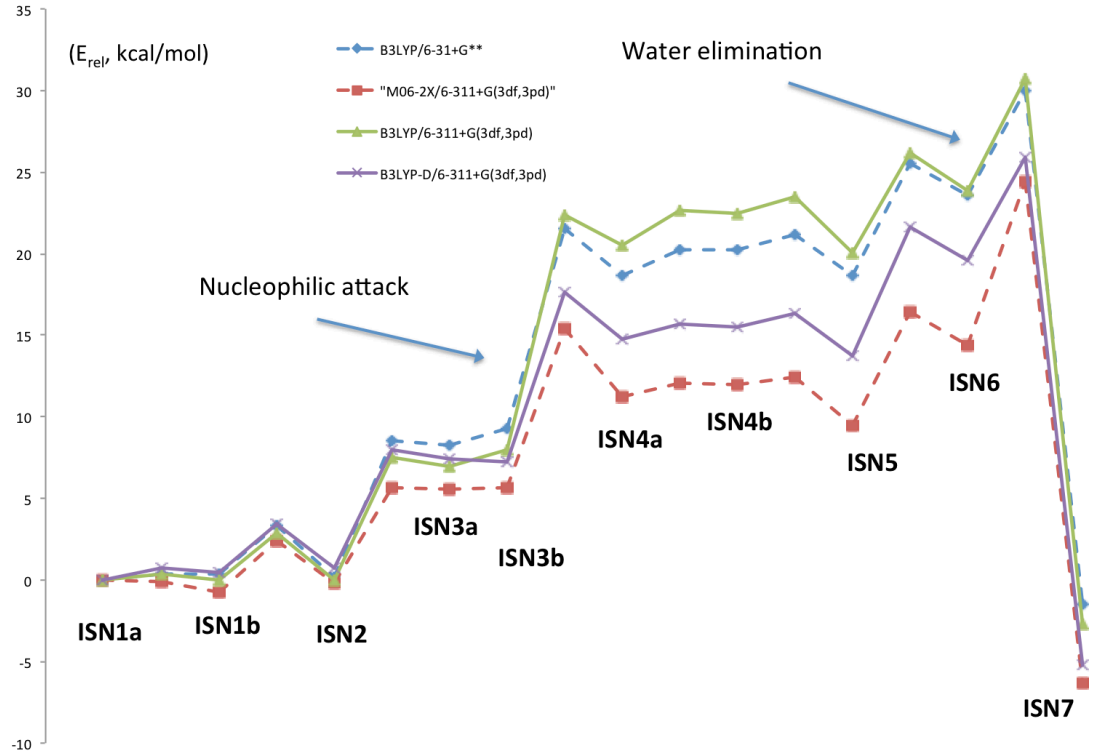


Figure 48 The dependence of different QM levels of theory on the QM/MM reaction profile.

Changing the DFT approximation, however, had a much larger impact on the energies of the reaction profile as shown in Figure 48. All relative energies of minima and saddle points following the nucleophilic attack are substantially lower at the M06-2X/6-311+G(3df,3pd) level and the highest point of the profile is now 24.4 kcal/mol. Suspecting inadequate description of dispersion interactions at the B3LYP level to be the reason for the difference in relative energies, we added a simple DFT-D dispersion

correction to all B3LYP energies.[26] Indeeed it turns out that the B3LYP-D profile is much closer to the M06-2X profile and these results suggest that it is necessary to take dispersion properly into account for reliable energies for our pathway. We note that M06-2X has been shown to be one of the best performing density functionals for general organic thermochemistry and barriers as was confirmed in a very extensive benchmarking study by Goerigk and Grimme.[34]

We were interested in exploring the dynamic behaviour of the system right after isopeptide bond formation, so we imported the QM/MM optimised **ISN7** structure into the CHARMM program, applied the CHARMM isopeptide bond topology to the the isopeptide bond (like in the intact isopeptide bond preparation in Chapter 5.2.2), redefined the reaction-formed water molecule as a TIP3 water molecule and started a classical MD simulation which ran for 1 ns. During the MD simulation the water molecule from the reaction moves from its QM/MM optimised position to a position where a water molecule is seen in the crystal structure. This is illustrated in Figure 49 which shows an overlay of the isopeptide bond region for the crystal structure, **ISN7** QM/MM optimised structure and the 1 ns snapshot from the MD simulation.

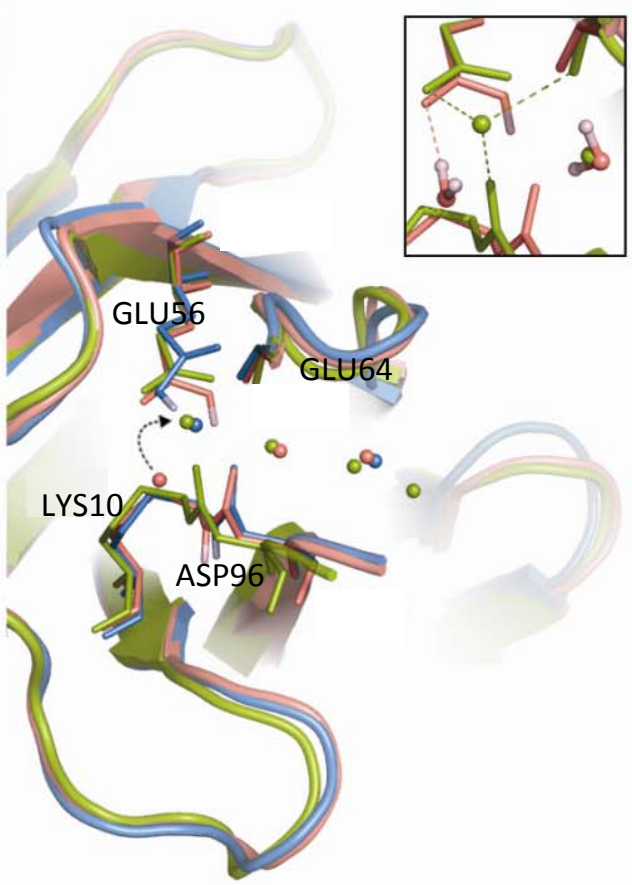


Figure 49 Overlay of the isopeptide region for the crystal structure (green), the **ISN7** QM/MM optimised structure (pink) and the last snapshot of the 1 ns MD simulation (cyan). Water molecules are shown as balls. The arrow indicates movement of the eliminated water to the position observed in both the MD and the crystal structure. Inset shows hydrogen bonds of this water molecule in these two locations.

This agreement between crystal structure and the computational structures for the position of the water molecule and the overall position of the residues in the isopeptide bond region as shown in Figure 49 is reassuring. The water molecule is stabilised by hydrogen bonds but we did discover when doing further MD simulations that this water molecule can escape from its position through a channel represented

in Figure 49 by 3 crystal structure water molecules that lead out into bulk solution. The eliminated water molecule thus is not trapped inside the protein but most likely is exchanged with water from bulk solution after reaction.

In QM/MM studies of potential energy profiles it is usually necessary to check for the sensitivity of the results to the conformational diversity of the protein. This involves using different snapshots from the classical protein equilibration and recalculating the same pathway a few times.[339] This enables one to check whether the initial snapshot chosen, was perhaps accidentally trapped in a high-energy local minimum that could lead to artifically low barriers. We recalculated our reaction profile using two additional snapshots (after 450 and 400 ps of classical MD simulation) which are shown in Figure 50.

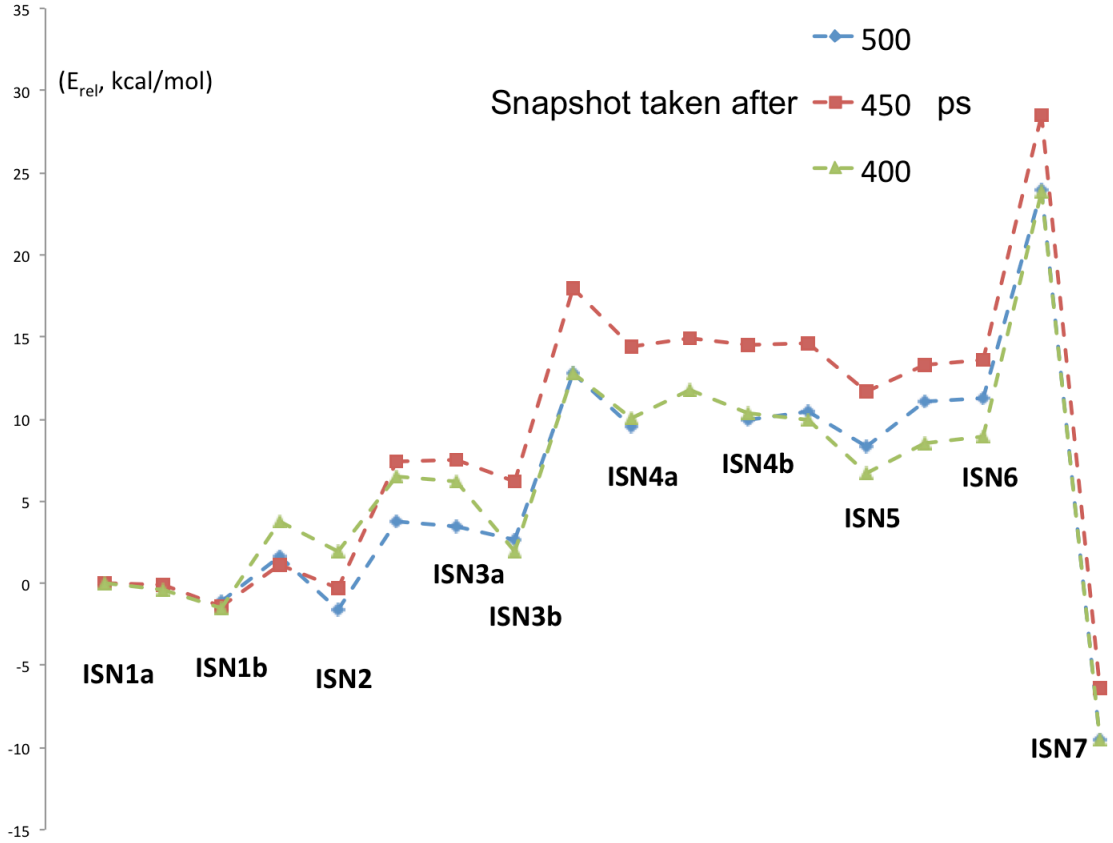


Figure 50 M06-2X/6-311+G(3df,3pd)/MM//B3LYP/6-31+G(d,p)/MM calculated reaction profiles starting from different snapshots of the classical MD equilibration.

Figure 50 shows that the 400 ps snapshot leads to a very similar reaction profile as the 500 ps snapshots, in particular the key barriers are of essentially the same height. The 450 ps snapshots leads to higher energies for many of the minima and transition states and particularly the two key transition states are considerably higher in energy. It is thus clear that the variability of the protein can affect the energetics of the calculated reaction profile and that protein variability for this system should be explored more. Ideally, this would be performed by direct free energy simulations of the protein that would sample the environmental degrees of freedom which should result in a single potential of mean force profile (e.g. by thermodynamic integration as in Chapter 4) , making multiple pathway calculations for many snapshots unnecessary. This remains very expensive, however, and was not attempted here.

The mechanism just described results in a surmountable activation barrier, due to active participation of the Glu56 residue in the nucleophilic attack step of Lys10 and Asp96 and the water dissociation step of the Lys10-Asp96 combined residue. The role of Glu56 in the nucleophilic attack step seems to be mainly due to the favourable hydrogen-bond network that it creates. This network and the proximity of the residue then allows convenient shuttling of protons through very low-energy barriers needed for the arrangement of residues for the water dissociation. Our mechanism suggests that a tetrahedral amide hydrate intermediate **ISN5** is a crucial low-energy intermediate for isopeptide bond formation to take place, rather than a concerted mechanism. Without the Glu56 residue it appears to be difficult to get to such a minimum. While the path to the zwitterionic tetrahedral intermediate **ISN4a** does not require any proton-shuttling through Glu56, this minimum seems to be stabilised by the hydrogen-bonds of Glu56 i.e. the minimum wouldn't be stable without the hydrogen bonding by Glu56 (see Chapter 5.4 for the mutant pathway where this is explored). The presence of Glu56 hence both stabilises the intermediate and allows the NH proton to be transferred away, following nucleophilic attack, which is necessary for subsequent water dissociation. The amide hydrate intermediate (e.g. **ISN5**) could be stabilised without Glu56 if the NH proton is transferred from the NH_2 group to the unprotonated oxygen on Asp96. This, however, would have to proceed through a very high-energy 4-center transition state and thus cannot occur. The water dissociation step could also occur without a tetrahedral intermediate and without Glu56 through a concerted mechanism but that would require the NH proton to be transferred to the Asp96 hydroxyl group at the same time as the nucleophilic attack step and which would also be a very high-energy transition state.

The highest barrier in our reaction profile is elimination of water from the neutral amide hydrate, **ISN6**. However, this may not be the rate-determining step. It is well known that water elimination of such diol-like intermediates occurs fast and is efficiently catalyzed by both acids and bases.[340] Since the isopeptide bond region is close to the bulk, it seems likely that water elimination could proceed even more efficiently by an acid- or base-catalyzed mechanism involving the solvent, which we did not explore. Thus we consider the key point of our mechanism to be the low barrier nucleophilic attack and formation of the tetrahedral intermediate.

5.4 Results: QM/MM calculations of Glu->Gln mutant of CnaB2

A mutation experiment where Glu56 was mutated into a glutamine residue (containing an amide group instead of a carboxylate group), and subsequent NMR analysis showed that the isopeptide bond had not been formed.[329] This suggests a crucial catalytic role for Glu56 which our proposed mechanism in Chapter 5.3 already seems to demonstrate.

While a GLN56 residue would be capable of forming hydrogen bonds like Glu56 it is not capable of acting as a proton shuttle (can neither donate or accept protons) and this single fact would seem to explain the catalytic role of Glu56 and why the reaction does not happen for the Glu56->Gln56 mutant.

Nonetheless, QM/MM calculations (B3LYP/6-31+G(d,p)/MM) were carried out for the GLN56 mutant where pathways for isopeptide bond formation were trialled,

involving the GLN56 residue as much as possible. A small 30 atom QM region was used.

A deprotonated Lys10 minimum, **GLNM1,** was found (See Figure 51) where Asp96 and GLN56 form a hydrogen-bonded amide-acid dimer similar to the acid dimer for the native protein. Nucleophilic attack by Lys10 was attempted, where the oxygen of the amide group acts as a hydrogen-bond acceptor to the incoming Lys10-NH_2 group. However, attempts to stabilise a tetrahedral intermediate, similar to the zwitterionic intermediate **ISN4a** for the native protein were not successful (falls back to **GLNM1**) , despite numerous attempts. This suggests that a carboxylic acid group is required for stabilisation of the tetrahedral intermediate leading to isopeptide bond formation and that a glutamate (or possibly an aspartate residue) is necessary for isopeptide bond formation involving Lys and Asp residues.

The only way to stabilise a tetrahedral intermediate for the mutant protein was to transfer the NH proton from Lys10 to the carbonyl oxygen of Asp96, making an amide hydrate tetrahedral intermediate, similar to **ISN5** of the native protein. However, this requires a proton transfer going through a 4-center transition state to occur simultaneously as the nucleophilic attack. This would not seem feasible and indeed the calculated barrier according to a test calculation turned out to be 37.5 kcal/mol at the M06-2X/6-311+G(3df,3pd)/MM level (single-points on B3LYP/6-31+G(d,p)/MM geometries). See Figure 51. Another 4-center transition state would then be required for water dissociation (since the proton-shuttle is not available) presumably requiring an energy penalty of similar magnitude. Another way to get isopeptide bond formation without stabilisation of the tetrahedral intermediate is through a concerted mechanism where the nucleophilic attack and water dissociation occur at the same time. This would, however, also require an unfavourable 4-center transition state.

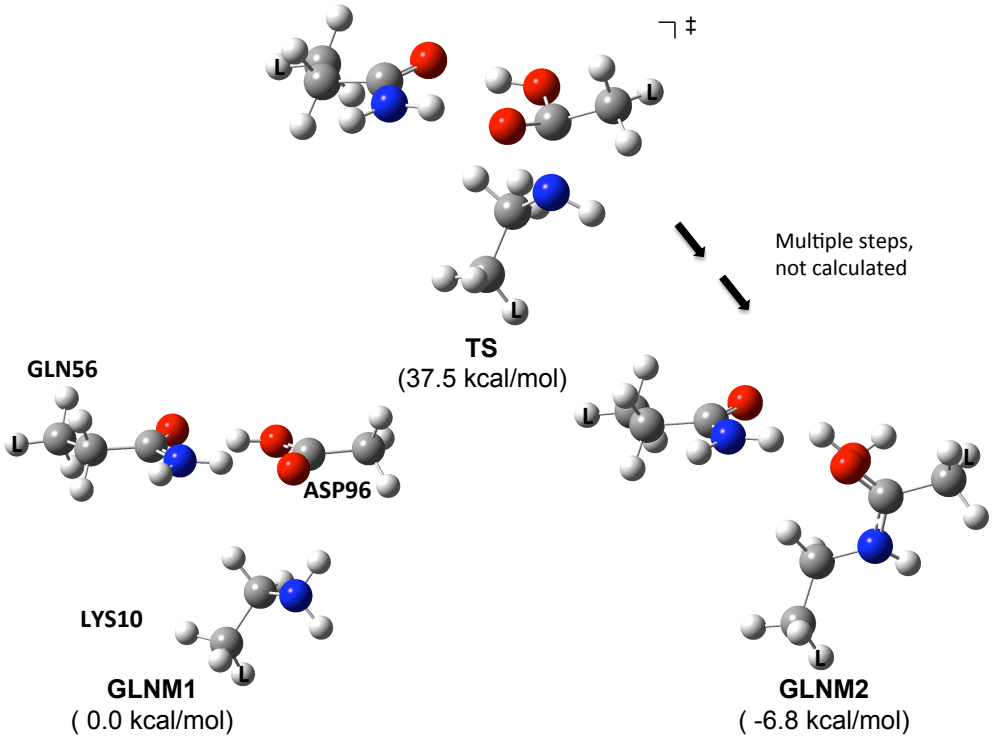


Figure 51 High-energy nucleophilic attack step leading to isopeptide bond formation for the GLN56 mutant. M06-2X/6-311+G(3df,3pd)/MM//B3LYP/6-31+G(d,p)/MM energies. Full QM region shown.

QM/MM calculations for the GLN56 mutant thus appear to confirm the basic mechanism we propose for the native protein; Glu56 acting as a favourable hydrogen-bonding partner for tetrahedral intermediate stabilisation and a proton shuttle. No low-energy isopeptide bond pathway seems possible for a mutant version without a carboxylate-containing sidechain for residue 56.

5.5 Results: Additional calculations using the frozen QM/MM preparation

While our proposed mechanism for Lys-Asp isopeptide bond formation in CnaB2 is generally satisfactory, experimental observations and speculations pose additional questions:

1. Is a *cis* isopeptide bond conformer possible?
2. Is an inverse mechanism possible: Lys10 reacting with Glu56 catalyzed by Asp96 ?
3. Why is isopeptide bond formation slower when Asp is mutated into Asn ?

In previous calculations we had also noticed that the distance between the nearby carbonyl group of the peptide bond of Glu64 and the Asp96 carboxyl group fluctuated a lot and was sometimes very close to the Asp96 carboxylic acid group. We hypothesised that the presence of this group could act as a hydrogen bond acceptor that would affect the energetics of the isopeptide reaction mechanism, potentially lowering the barrier for water dissociation and increase the stability of the tetrahedral amide hydrate intermediates. Using the geometry of the amide hydrate intermediate just before water dissociation (equivalent to **ISN6** in Figure 47) we then carried out a classical MD simulation keeping the Lys10, Asp96 and Glu56 sidechains fixed (frozen QM/MM preparation). During the 1 ns simulation a hydrogen bond between the Asp96 hydroxyl group and the carbonyl group of Glu64 formed (H-bond distance of ~1.9 Å). The 1 ns snapshot was then used for QM/MM calculations.

In order to perform faster calculations the RI-B97D functional was used instead of the B3LYP functional. Additionally a slightly smaller QM region of 56 atoms was used. These changes to the computational protocol unfortunately makes comparison with the previous calculations slightly problematic although it seems unlikely that B3LYP and RI-B97D would predict very different geometries or that the slightly smaller QM region would affect the overall energetics (single-point energy evaluations are performed at the M06-2X/6-311+G(3df,3pd) level).

The QM/MM pathway calculations were performed, starting from the amide hydrate intermediate and the product and reactant minima located. However, only the transition states for the nucleophilic attack and water dissociation were located and used to calculate relative energies with respect to a cleaved reactant state as shown in Figure 52. Curiously, the calculated barriers towards nucleophilic attack and water dissociation are rather different this time around with the relative energy of the water dissociation step being significantly lower than previously at 15.5 kcal/mol while the relative energy of the nucleophilic attack step is higher at 18.9 kcal/mol.

This then suggests that the barrier to water dissociation is more easily surmountable than previous calculations indicated and that the nearest environment can influence this barrier (we also note that water molecules from the bulk are nearby that could also influence the mechanism by forming hydrogen bonds in a similar way or act as acid/base catalysts).

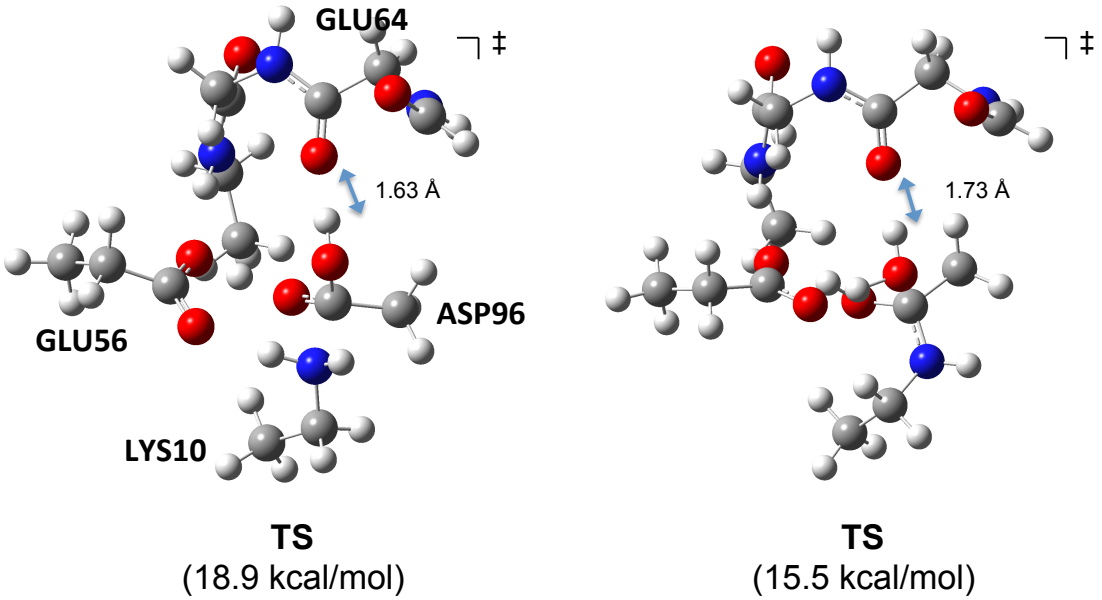


Figure 52 The nucleophilic attack transition state (left) and the water dissociation transition state (right), when calculated using the frozen QM/MM preparation. Hydrogen bonds to the Glu64 carbonyl group are indicated. M06-2X/6-311+G(3df,3pd)/MM//RI-B97-D/6-31+G(d,p)/MM relative energies w.r.t. to the reactant state.

It is not clear why the nucleophilic attack step becomes slightly less favourable than before. It may be that the Glu64 residue is in too close contact with the Asp96 residue to be favourable (due to oxygen-oxygen repulsions). How much the key barriers are affected by nearby protein residues like Glu64 is hence an open question, possibly best explored by free energy simulations.

We also performed calculations for this setup involving an isopeptide bond with a *cis* conformation instead of trans as Lys-Asn *cis* isopeptide bonds have been found to be as common as trans.[324] Despite the double hydrogen bond to Glu56 in the *cis* conformation, the *trans* conformer is considerably more stable by 14.2 kcal/mol (see Figure 53). The whole pathway to *cis* bond formation was not calculated but should be quite similar as the pathway to *trans* isopeptide bond formation.

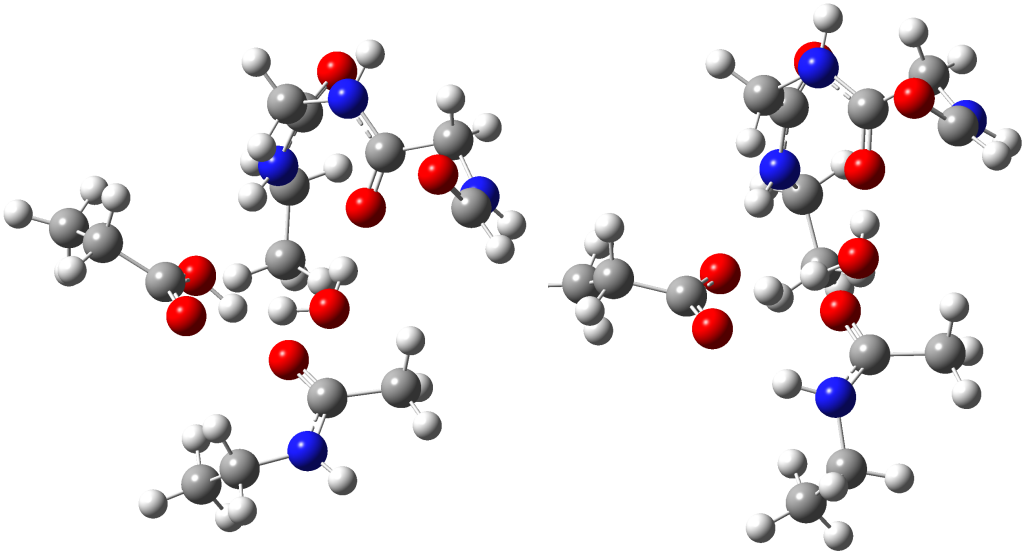


Figure 53 *Trans* vs. *cis* isopeptide bonds. M06-2X/6-311+G(3df,3pd)/MM//RI-B97-D/6-31+G(d,p)/MM relative energies with respect to the reactant state.

Our calculations thus predict that *trans* isopeptide bonds are thermodynamically much more favourable than *cis* isopeptide bonds for our system (not unexpectedly) despite the increased hydrogen bonding of the *cis* conformer. We did not calculate the barrier to *cis* isopeptide bond formation, however, it is clear that *cis* bond formation is just going to involve a slight shift of the OCNH dihedral of the transition state shown in Figure 53 (right), perhaps slightly increasing/decreasing the barrier height.

We note that Lys-Asn isopeptide bonds have been found in both *cis* and *trans* conformations in bacterial proteins. The only other case of a Lys-Asp isopeptide bond was recently found in a *S. pyogenes* pilus protein having both a Lys-Asp and a Lys-Asn isopeptide bond.[341] It is striking that the Lys-Asp isopeptide bond in that protein is *trans* while the Lys-Asn bond is cis. While it may be that the environment of the proteins containing Lys-Asn isopeptide bonds makes *cis* isopeptide bonds thermodynamically more favourable it seems unlikely as Lys-Asp and Lys-Asn isopeptide bonds are identical in chemical composition and the environment is usually very similar (hydrophobic and a nearby glutamate). We thus hypothesise that the *cis* isopeptide bond may simply be the slightly more kinetically favoured product in Lys-Asn proteins. In Lys-Asp proteins, the kinetics of *cis* vs. *trans* formation may be similar; however, since the *cis* product is so thermodynamically unfavourable in the Lys-Asp case, a reverse reaction might easily occur (water re-attacking the isopeptide bond), creating an equilibrium that would eventually result in the formation of the most favourable product, the *trans* isopeptide bond. A detailed mechanistic study of *cis* vs. *trans* isopeptide bonds for both Lys-Asp and Lys-Asn proteins is needed to shed further light on these issues.

An additional question that our proposed mechanism of Lys-Asp bond formation in the CnaB2 system creates, is the possibility of an inverse reaction occurring, where instead of Lys10 reacting with Asp96 to form an isopeptide bond catalyzed by Glu56; Lys10 instead reacts with Glu56 catalyzed by Asp96. Based on our proposed

mechanism, there is nothing to suggest that this could not happen, as both Glu56 and Asp96 have carboxylic acid sidechains. This possibility needs to be explored.

Yang and coworkers recently published a QM/MM mechanism of a Lys-Asn *cis* isopeptide bond formation in the *S. pyogenes* Spy0128 pilus protein by QM/MM free energy simulations.[342] This remains the only other reported calculated mechanism of isopeptide bond formation to date, but for a different protein and a slightly different reaction. The authors similarly came to the conclusion that the lysine must be deprotonated before nucleophilic attack and that a tetrahedral intermediate is formed and suggested a very similar purpose of the catalytic glumatate residue as for our protein, the only difference being in how the proton transfer steps are carried out which should be of minor importance.

Another open question is the slower isopeptide bond formation of the Asp->Asn mutant. We did some preliminary explorations of this issue by taking the frozen QM/MM setup, replaced one of the OH groups (the one later H-bonding to Glu64) with NH_3 , re-defined the topology, and re-equilibrated with most of the atoms of Lys10, Asn96 and Glu56 frozen as before. We then explored the formation of Lys-Asn isopeptide bond formation for the 1 ns snapshot, using the RI-B97D and M06-2X functionals and the same basis sets as before.

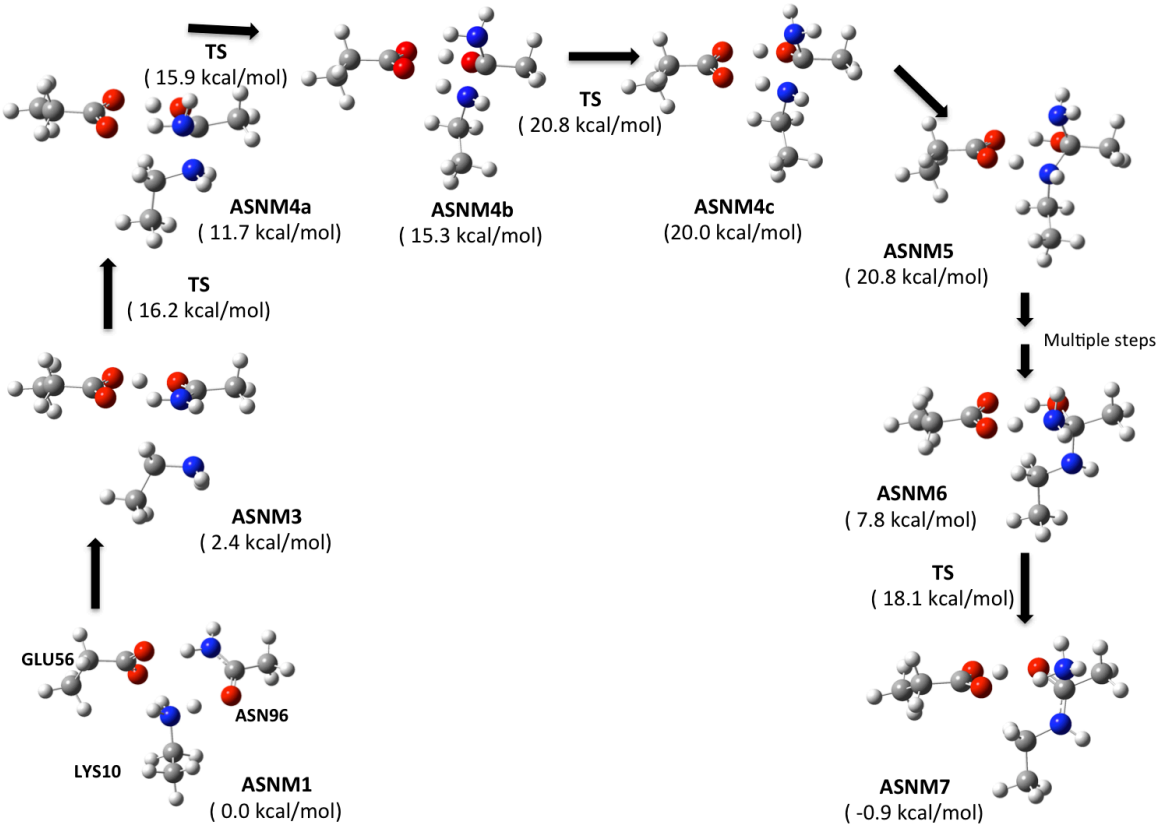


Figure 54. Preliminary reaction mechanism for the Asp->Asn mutant. M06-2X/6-311+G(3df,3pd)/MM//RI-B97-D/6-31+G(d,p)/MM relative energies. Not all intermediates shown and not all transition states have been calculated.

The preliminary results for this pathway are shown in Figure 54. The reaction proceeds similarly by a nucleophilic attack, lower in energy than the nucleophilic

attack in Figure 52, but then has to go through a few higher energy re-arrangement steps, before ammonia elimination. These re-arrangement steps involve moving the NH_2 group around to better stabilise the amino-alcohol intermediate and set the system up for ammonia elimination. It seems that due to the "bulkiness" of the amino group, these re-arrangements may be slightly more repulsive than for the hydroxyl group. However, this mechanism needs to better characterised and reaction profiles for many more MD snapshots for this mutant and the native enzyme would have to be calculated in order to properly assess if this "bulkiness" of the amino group is the difference between slower isopeptide bond formation in the Asn96 mutant.

5.6 Results: The isopeptide bond mechanism in the gas-phase

Modelling of intramolecular protein reactions such as isopeptide bond formation is made convenient by the QM/MM methodology. In order to gauge the effect of the protein environment, gas-phase modelling of the basic reaction was undertaken.

Specifically we were interested if an amidation reaction calculated using a part of the previously employed QM region in the gas-phase, would be favourable.

The reaction profile was calculated at the B3LYP/6-31+G(d,p) level with single-point energy calculations at the M06-2X/6-311+G(3df,3pd) level, i.e. the same level of theory as in the QM/MM calculations. The Synchronous Transit and Quasi-Newton method (STQN)[343] in Gaussian was used to locate transition states (QST2 and QST3 keywords) all of which were confirmed by frequency calculations which were also used to calculate the thermal correction to enthalpy and free energy.

The computed potential energy profile ΔE (0K, no ZPVE), enthalpy profile ΔH (298 K) and free energy profile ΔG (298 K) are shown in Figure 55.

The gas-phase mechanism is slightly different than the QM/MM calculated mechanism. First of all, an NH_3^+ group on the lysine-like residue is not stable and leads to spontaneous proton transfer to the carboxylic acid groups and no zwitterionic intermediates are found on the reaction profile unlike the QM/MM profile. Apart from these differences, a very similar reaction profile is predicted, a similar nucleophilic attack step, proton shuttling involving the glutamate-model and water dissociation. Transition states between **GAS1** & **GAS3** and **GAS6** & **GAS7** were troublesome to locate with STQN algorithms (subtle conformational changes of the reactant and product states) and were eventually abandoned as they are not very important for the mechanism.

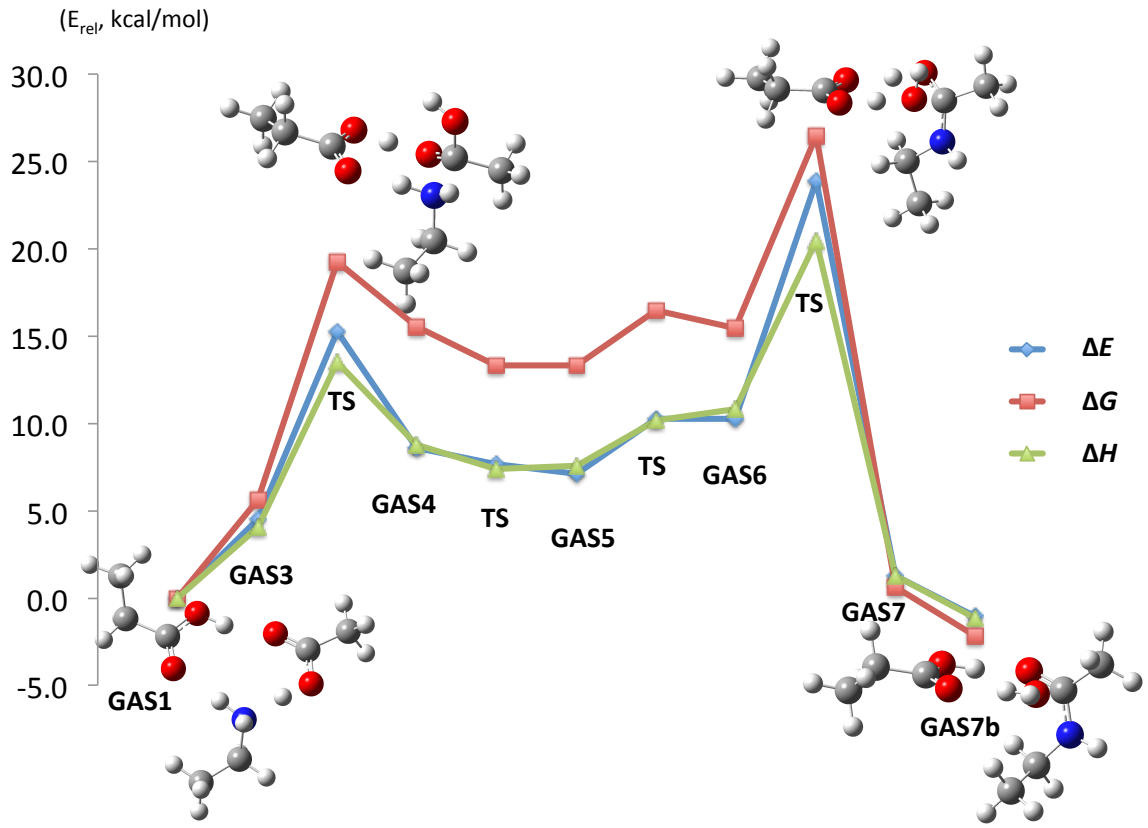


Figure 55 The gas-phase reaction profile calculated at the M06-2X/6-311+G(3df,3pd)//B3LYP/6-31+G(d,p) level. Potential energy profile and free energy profile shown (free energy corrections at the B3LYP/6-31+G(d,p) level. The numbering of minima was deliberately made similar as the numbering in the QM/MM profile.

The energetics of the 0 K potential energy pathway are very similar to the original QM/MM pathway shown in Figure 45 and 48. Interestingly, the Δ*H* pathway is for the most part very similar as the Δ*E* pathway, but the energies of the two main barriers are lowered, presumably due to favourable loss of vibrational energy of the bonds being broken in the transition states. However, the 298 K free energy pathway raises the barriers substantially, due to unfavourable entropy effects, which would presumably prevent reactions of this kind from occurring spontaneously in the gas-phase or solution. We do note, however, that the accuracy of the free energy correction is unknown. Due to the many low-frequency vibrations of this system, harmonic vibrational frequencies may be inaccurate for the entropy contribution.[1]

Free energy simulations of the QM/MM isopeptide bond pathway were not attempted as they require lengthy MD simulations so it is currently not known whether a calculated QM/MM free energy profile of the mechanism would raise the barriers considerably. It may be that isopeptide bond formation occurs primarily because the entropy penalty has already been paid by how the residues are positioned inside the protein, i.e. it is the folding of the peptide that brings the residues together and allows the reaction to occur in the first place. Recent studies of free-energy effects in enzymatic reactions suggest that for the most part, entropy effects are low and that free-energy profiles compare well to potential energy profiles as the entropy penalty of binding and pre-organisation of the substrates has already been paid.[344,345] However, it may also be that calculation of the ZPVE and thermal corrections to enthalpy would lower the barriers in our QM/MM mechanisms of isopeptide bond formation.

Finally, we note that our computed mechanism of amide bond formation in the gas-phase is intriguing, despite the high free energy barriers. Amide bond formation has been described as one of the most important reactions in organic chemistry and amide bonds are common in drug molecules and biologically relevant compounds. Current synthetic methods to create amide bonds have come under scrutiny due to waste and expense and it is clear that new and better methods to create amide bonds are needed, especially for the growing area of synthesis and modification of peptides.[346] The gas-phase mechanism in Figure 55 would only be applicable as a catalytic strategy if it is favourable to bring the 3 molecules together in solution, i.e. if the intermolecular forces of the **GAS1** structure in solution (or a similar complex) are stronger than the associated entropic penalty. Calculations suggest this to be the case in the gas-phase by ~ 3 kcal/mol (corrected for basis set superposition error) but this may not be the case in solution. Some kind of supramolecular complex might be required to accomplish this in practice (by somehow trapping the substrates and bringing them close together).[347]

5.7 Summary

A mechanism of the recently discovered Lys-Asp isopeptide bond in a surface bacterial protein has been computed by QM/MM calculations. The mechanism explains some key experimental observations such as the catalytic role of the glutamate residue and the reason for the experimental observation that no isopeptide bond is formed in mutants with no glutamate.

There are still a number of open questions for spontaneous formation of isopeptide bonds in bacterial proteins, such as whether Lys-Asn and Lys-Asp isopeptide bonds are formed by a common mechanism, why the Lys-Asp bond is formed more rapidly than the Lys-Asn bond in the CnaB2 mutant, the preference for *cis* vs. *trans* isopeptide bonds and many others.

We suggest future computational studies of isopeptide bond formation to utilise molecular dynamics and free energy simulations in a QM/MM scheme. Molecular dynamics simulation should enable one to better understand the importance of the environment on the mechanism (our calculations already suggest that such effects could stabilise the water dissociation step considerably), check for the flexibility of the protein and interaction of the reaction site with the nearby bulk solvent as well as giving a clearer picture of how entropy affects the kinetics of the reaction. Such MD simulations would most likely require the use of semi-empirical methods due to computational cost and we note the successful use of the PM3 method in the 3F-GABA study. The recent OMX methods have also been shown to be remarkably successful in a recent benchmark study of organic thermochemistry, kinetics and weak interactions.[321] OMX/MM free energy simulations could be performed either using thermodynamic integration or umbrella sampling to yield a free energy profile and could possibly be corrected to a DFT/MM energy profile by using a thermodynamic cycle as in the 3F-GABA study. The recent use of QM/MM FEP methods on frozen NEB pathways also looks promising.[344,345]

Finally we note that nature has of course already come up with an elegant way of creating amide bonds in the peptidyltransferase centre in the ribosome where the

peptidyl-tRNA residue is transferred to the aminoacyl-tRNA residue and the peptide bond is created. The mechanism of peptide bond mechanism is still under debate.[348-355] The main rate-enhancing effect of peptide bond formation in the ribosome has been suggested to be entropic in origin,[348] that is achieved by desolvation and positioning of the substrates, leading to a small TΔS factor. However, such observations do not explain the mechanism (whether catalytic or not). Several studies have suggested that the peptidyl-tRNA 2'-OH is involved and acts as a proton shuttle leading to possible 6- to 8-membered transition states[349-352] and tetrahedral intermediates, but the importance of the 2'-OH group has been challenged as well.[353] Most recently, kinetic isotope analyses[354] have ruled out a completely concerted mechanism and suggest instead a stepwise mechanism where the C-N bond formation (nucleophilic attack by the aminoacyl group) leading to a tetrahedral intermediate is rate-limiting. A proton-shuttle role by peptidyl-tRNA 2'-OH is still conceivable but its role might also just be that of a stabilising hydrogen bond (possibly involved in the orientation of the substrate for subsequent nucleophilic attack) while nearby water molecules are responsible for the necessary proton transfers (deprotonation of the aminoacyl-tRNA amine group and protonation of the peptidyl-tRNA 3'O leaving group).

The currently proposed ideas for peptide bond formation in the ribosome are thus not too dissimilar to the proposed mechanisms discussed in our work on isopeptide bond formation where orientation of the residues for nucleophilic attack, stabilisation of a tetrahedral intermediate and low barrier proton transfers are all key points and may well be nature's efficient way of making amide bonds.

Conclusion

We have presented several computational studies that revolve around accounting for environmental effects in the modelling of chemical systems. For the most part we have used the quantum mechanics / molecular mechanics (QM/MM) methodology in different ways that have allowed us to model complex chemical systems of very different size and form.

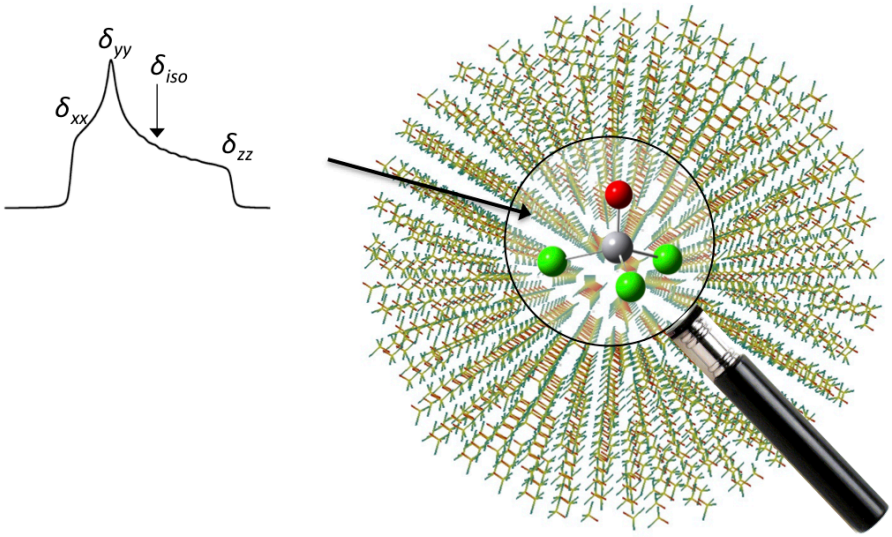


Figure 56 Modelling solid-state NMR spectra of molecular crystals (Chapter 2).

The main project concerned first principles computations of solid-state NMR properties. As well as studying chemical shifts and quadrupole couplings in the gas phase for validation of the electronic structure methodology, we devised a general QM/MM based protocol to model local geometries as well as spectroscopic properties for essentially any molecule in the solid state, taking into account solid-state geometrical effects and direct effects on the molecular properties. The molecule of interest is described at an appropriate QM level and is embedded in a large array of MM atoms built up from crystal structure information. The MM atoms are described by point charges and short-range Lennard-Jones potentials. The point charges are derived from a single molecule QM calculation and are iterated in the field of point charges. The protocol allows one to perform local geometry optimisation and embedded molecular property calculations of essentially any molecule. We have applied this protocol to many different systems both for validation as well as insight into how the solid-state affects molecular properties: e.g. the molecular crystal of HCN-BF_3 with its dramatic gas-to-solid bond contraction, the solid-state NMR properties of vanadium compounds and the solid-state ^{59}Co properties of the ionic cobaloxime complex and the wet crystal of vitamin B_{12}.

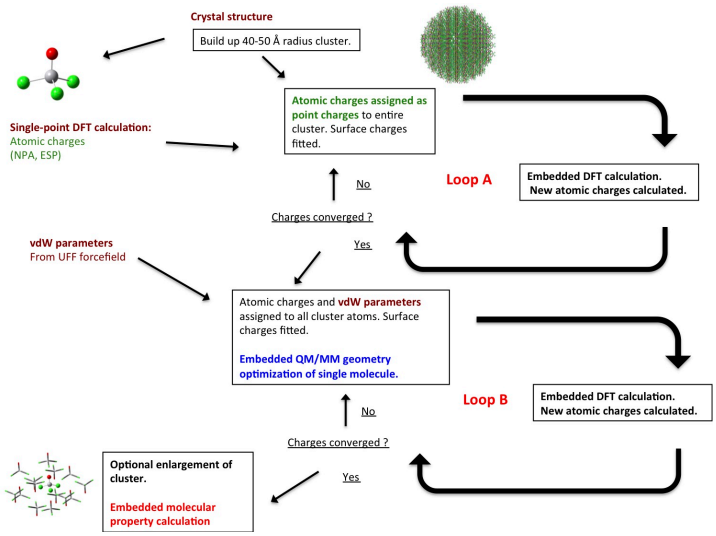


Figure 57 The QM/MM protocol for modelling local properties of molecular crystals (Chapter 2).

In a second project we have shed light on the complicated conformational properties of the 3F-GABA zwitterion in aqueous solution.

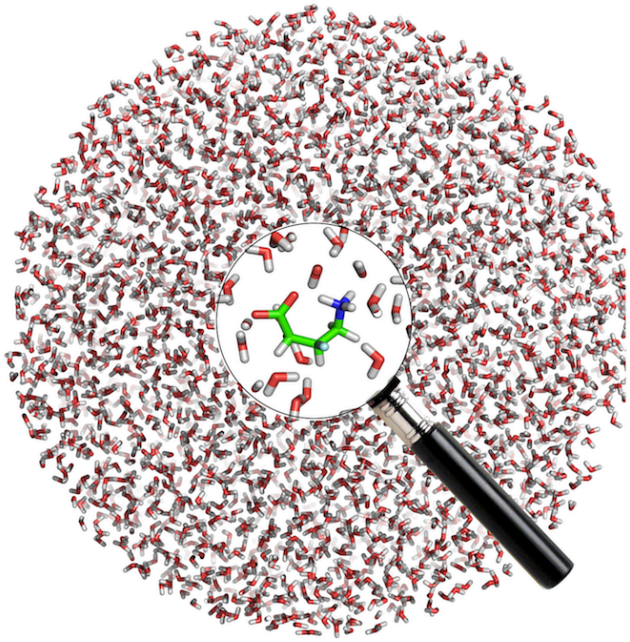


Figure 58 Solvation of the flexible zwitterion 3F-GABA in a 6000 water molecule QM/MM cluster (Chapter 3).

A previous computational analysis using continuum solvation models was not able to describe the conformational properties correctly and showed bias towards folded conformers. Using elaborate QM/MM molecular dynamics simulations of 3F-GABA in an explicit large water cluster (see Figure 58), we computed free energy differences that revealed that extended conformers are slightly favored in solution in agreement with the experimental NMR analysis. A spin-spin coupling constant analysis using dynamically averaged data that took the solvation effects into account as well, reaffirmed the original experimental NMR analysis.

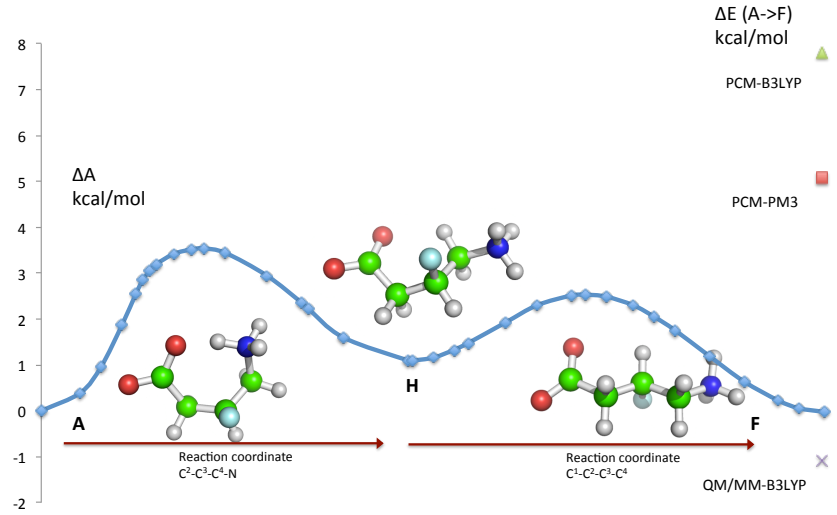


Figure 59 The computed PM3/MM free energy profile from conformer **A** to **F** (Chapter 3).

In a third project we have contributed to the mechanistic understanding of spontaneous intramolecular isopeptide bond formation in bacterial peptides. Isopeptide bonds are the result of inter- & intra-molecular amidation reactions between sidechains of amino acids of proteins and have recently been found to be widely occurring in surface bacterial proteins. The first example of an isopeptide bond between a lysine and aspartate sidechains was recently discovered by collaborators at the University of St Andrews.

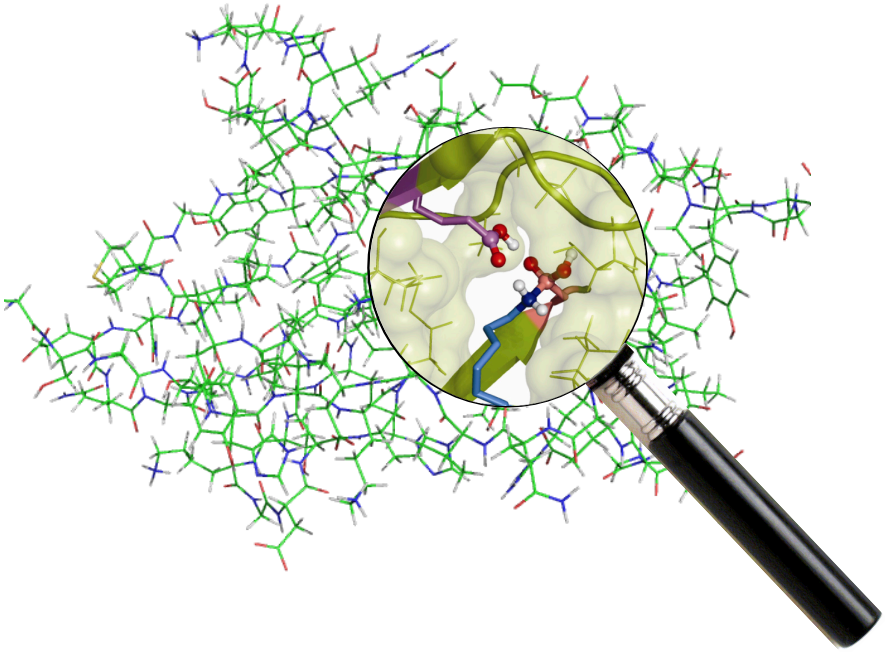


Figure 60 Modelling the intramolecular isopeptide bond formation in a bacterial peptide (Chapter 4).

In collaboration with the experimental scientists we have uncovered the mechanistic aspects of isopeptide bond formation that explain how and why this reaction takes place inside the protein. A QM/MM calculated reaction profile reveals the crucial catalytic role that a nearby glutamate residue has, acting as an efficient proton shuttle

and hydrogen-bonding residue that leads to low-barrier nucleophilic attack of lysine towards aspartate, stabilisation of a tetrahedral amide hydrate intermediate and subsequent water elimination.

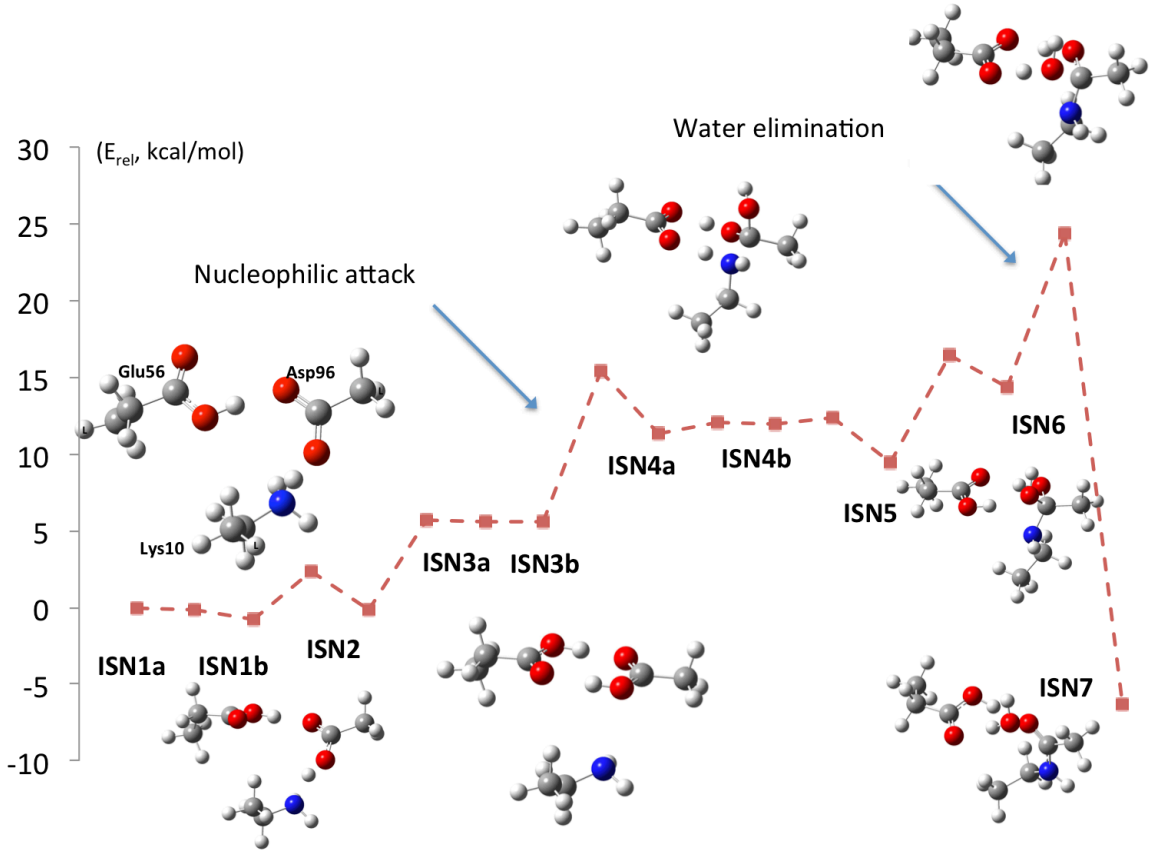


Figure 61 The isopeptide bond reaction mechanism (Chapter 4).

References

(1) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; 2nd ed. Wiley, 2004; pp. 105–114,122–129,167,309–315,331–338,375–377,394–402,429–437.
(2) Jensen, F. *Introduction to computational chemistry*; Wiley: Wiltshire, 2007; pp. 115–131, 277–292, 451–452.
(3) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.
(4) Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
(5) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2006**, *117*, 185–199.
(6) Bernstein, N.; Kermode, J. R.; Csányi, G. *Rep. Prog. Phys.* **2009**, *72*, 026501.
(7) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory, 2nd Edition*; 2nd ed. Wiley-VCH: Weinheim, 2001; pp. 3–12,19.
(8) Neese, F. *Coord. Chem. Rev.* **2009**, *253*, 526–563.
(9) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, 864–871.
(10) Thomas, L. H. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542–548.
(11) Fermi, E. *Rend. Accad. Lincei* **1927**, *6*, 602–607.
(12) Xia, J.; Huang, C.; Shin, I.; Carter, E. A. *J. Chem. Phys.* **2012**, *136*, 084102.
(13) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133–1138.
(14) Slater, J. C. *The Self-Consistent Field for Molecular and Solids, Quantum Theory of Molecular and Solids*; McGraw-Hill: New York, 1974; Vol. 4.
(15) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
(16) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244–13249.
(17) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
(18) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
(19) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.
(20) Scuseria, G. E.; Staroverov, V. N. In *Theory and Applications of Computational Chemistry — The First Forty Years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds. Elsevier, 2005; pp. 669–724.
(21) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(22) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
(23) Henderson, T. M.; Janesko, B. G.; Scuseria, G. E. *J. Phys. Chem. A* **2008**, *112*, 12530–12542.
(24) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215–241.
(25) Grimme, S. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.
(26) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
(27) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(28) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*, 244103.
(29) Hujo, W.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 3866–3871.
(30) Kozuch, S.; Martin, J. M. L. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104–20107.
(31) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.
(32) Liakos, D. G.; Neese, F. *J. Chem. Theory Comput.* **2011**, *7*, 1511–1523.
(33) Perdew, J. P.; Schmidt, K. *AIP Conference Proceedings* **2001**, *577*, 1–20.
(34) Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.
(35) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(36) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220.
(37) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.

(38) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724–728.
(39) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
(40) Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acc.* **1973**, *28*, 213–222.
(41) Clark, T.; Chandrasekhar, J.; Spitznagel, G. N. W.; Schleyer, P. V. R. *J. Comput. Chem.* **1983**, *4*, 294–301.
(42) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265–3269.
(43) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033–1036.
(44) Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377–4384.
(45) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
(46) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
(47) Wilson, A. K.; van Mourik, T.; Dunning, T. H., Jr *J. Mol. Struct.: THEOCHEM* **1996**, *388*, 339–349.
(48) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
(49) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.
(50) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
(51) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(52) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
(53) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
(54) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737–7743.
(55) Otter, den, W. K. *J. Chem. Phys.* **2000**, *112*, 7283.
(56) Laidler, K. J.; King, M. C. *J. Phys. Chem.* **1983**, *87*, 2657–2664.
(57) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771–12800.
(58) Fernández-Ramos, A.; Miller, J. A.; Klippenstein, S. J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 4518–4584.
(59) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107–115.
(60) Neese, F. In *High Resolution EPR*; Berliner, L.; Hanson, G., Eds. Springer New York: New York, NY, 2009; Vol. 28, pp. 175–229.
(61) Schwerdtfeger, P.; Pernpointner, M.; Nazarewicz, W. In *Calculation of NMR and EPR Parameters: Theory and Applications*; Kaupp, M.; Bühl, M.; Malkin, V. G., Eds. Wiley-VCH: Weinheim, 2004; pp. 279–291.
(62) Adiga, S.; Aebi, D.; Bryce, D. L. *Can. J. Chem.* **2007**, *85*, 496–505.
(63) Gaussian 03, Revision E.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, 2004.
(64) Gaussian 09, Revision A.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.;

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

(65) Gauss, J.; Stanton, J. F. In *Calculation of NMR and EPR Parameters: Theory and Applications*; Kaupp, M.; Bühl, M.; Malkin, V. G., Eds. Wiley-VCH: Weinheim, 2004; pp. 123–139.

(66) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789–807.

(67) Haeberlen, U. In *Advances in Magnetic Resonance; Suppl. 1*; Waugh, J. S., Ed. Academic Press: New York, 1976.

(68) Mehring, M. *Principles of High Resolution NMR in Solids, 2nd. ed.*; Springer Verlag: Berlin, 1983.

(69) Spiess, H. W. In *NMR Basic Principles and Progress*; Diehl, P.; Fluck, E.; Kosfeld, R., Eds. Springer Verlag: Berlin, 1978; Vol. 15.

(70) Cremer, D.; Gräfenstein, J. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2791–2816.

(71) Ramsey, N. *Phys. Rev.* **1953**, *91*, 303–307.

(72) Jensen, F. *J. Chem. Theory Comput.* **2008**, *4*, 719–727.

(73) Jensen, F. *J. Chem. Theory Comput.* **2006**, *2*, 1360–1369.

(74) Wiberg, K. B.; Rablen, P. R. *J. Comput. Chem.* **1993**, *14*, 1504–1518.

(75) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.

(76) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.

(77) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(78) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(79) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098–16104.

(80) Senn, H. M.; Thiel, W. In *Topics in Current Chemistry*; Reiher, M., Ed. Springer Berlin Heidelberg, 2007; Vol. 268, pp. 173–290.

(81) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(82) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(83) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

(84) Wesolowski, T. A. In *Computational Chemistry: Reviews of Current Trends*; World Scientific, 2006; Vol. 10, pp. 1–82.

(85) Jacob, C. R.; Neugebauer, J.; Visscher, L. *J. Comput. Chem.* **2008**, *29*, 1011–1018.

(86) Jacob, C. R.; Neugebauer, J.; Jensen, L.; Visscher, L. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2349–2359.
(87) Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed. NIC Proceedings: Jülich, 2000; Vol. 3, pp. 285–305.
(88) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1–28.
(89) ChemShell, a Computational Chemistry Shell, see www.chemshell.org.
(90) Philipsborn, von, W. *Chem. Soc. Rev.* **1999**, *28*, 95–105.
(91) Ashbrook, S. E. *Phys. Chem. Chem. Phys.* **2009**, *11*, 6892–6905.
(92) Ronconi, L.; Sadler, P. J. *Coord. Chem. Rev.* **2008**, *252*, 2239–2277.
(93) Murray, R. K.; Granner, D. K.; Harper, H. A.; Rodwell, V. W. *Harper's Illustrated Biochemistry*; McGraw-Hill Medical, 2003.
(94) Pooransingh-Margolis, N.; Renirie, R.; Hasan, Z.; Wever, R.; Vega, A. J.; Polenova, T. *J. Am. Chem. Soc.* **2006**, *128*, 5190–5208.
(95) Lipton, A. S.; Heck, R. W.; Ellis, P. D. *J. Am. Chem. Soc.* **2004**, *126*, 4735–4739.
(96) Lipton, A. S.; Heck, R. W.; Primak, S.; McNeill, D. R.; Wilson, D. M.; Ellis, P. D. *J. Am. Chem. Soc.* **2008**, *130*, 9332–9341.
(97) Lipton, A. S.; Heck, R. W.; de Jong, W. A.; Gao, A. R.; Wu, X.; Roehrich, A.; Harbison, G. S.; Ellis, P. D. *J. Am. Chem. Soc.* **2009**, *131*, 13992–13999.
(98) Frenking, G. In *Encyclopedia of Computational Chemistry*; R Schleyer, von, P., Ed. Wiley: New York, 1998; pp. 3073–3084.
(99) Hyla-Kryspin, I.; Koch, J.; Gleiter, R.; Klettke, T.; Walther, D. *Organometallics* **1998**, *17*, 4724–4733.
(100) Hyla-Kryspin, I.; Grimme, S. *Organometallics* **2004**, *23*, 5581–5592.
(101) Bühl, M. *Ann. Rep. NMR Spectrosc.* **2008**, *64*, 77–126.
(102) Bühl, M.; Parrinello, M. *Chem. Eur. J.* **2001**, *7*, 4487–4494.
(103) Bühl, M. *J. Phys. Chem. A* **2002**, *106*, 10505–10509.
(104) Bühl, M.; Mauschick, F. T. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5508–5514.
(105) Bühl, M.; Grigoleit, S.; Kabrede, H.; Mauschick, F. T. *Chem. Eur. J.* **2005**, *12*, 477–488.
(106) Bühl, M.; Wrackmeyer, B. *Magn. Reson. Chem.* **2010**, *48*, S61–S68.
(107) Bochevarov, A. D.; Friesner, R. A.; Lippard, S. J. *J. Chem. Theory Comput.* **2010**, *6*, 3735–3749.
(108) Sandala, G. M.; Hopmann, K. H.; Ghosh, A.; Noodleman, L. *J. Chem. Theory Comput.* **2011**, *7*, 3232–3247.
(109) Chatterjee, P. B.; Goncharov-Zapata, O.; Quinn, L. L.; Hou, G.; Hamaed, H.; Schurko, R. W.; Polenova, T.; Crans, D. C. *Inorg. Chem.* **2011**, *50*, 9794–9803.
(110) Cuny, J.; Sykina, K.; Fontaine, B.; Le Polles, L.; Pickard, C. J.; Gautier, R. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19471–19479.
(111) Stueber, D. *Concepts Magn. Reson.* **2006**, *28A*, 347–368.
(112) Stueber, D.; Guenneau, F. N.; Grant, D. M. *J. Chem. Phys.* **2001**, *114*, 9236–9243.
(113) Stueber, D.; Grant, D. M. *J. Am. Chem. Soc.* **2002**, *124*, 10539–10551.
(114) Stueber, D.; Grant, D. M. *Solid State Nucl. Magn. Reson.* **2002**, *22*, 439–457.
(115) Orendt, A. M. *Magn. Reson. Chem.* **2006**, *44*, 385–389.
(116) Götz, K.; Meier, F.; Gatti, C.; Burow, A. M.; Sierka, M.; Sauer, J.; Kaupp, M. *J. Comput. Chem.* **2010**, *31*, 2568–2576.
(117) Pickard, C.; Mauri, F. *Phys. Rev. B* **2001**, *63*, 245101.

(118) Yates, J.; Pickard, C.; Mauri, F. *Phys. Rev. B* **2007**, *76*, 024401.
(119) Sebastiani, D.; Parrinello, M. *J. Phys. Chem. A* **2001**, *105*, 1951–1958.
(120) Skachkov, D.; Krykunov, M.; Kadantsev, E.; Ziegler, T. *J. Chem. Theory Comput.* **2010**, *6*, 1650–1659.
(121) Skachkov, D.; Krykunov, M.; Ziegler, T. *Can. J. Chem.* **2011**, *89*, 1150–1161.
(122) Blöchl, P. E. *Phys. Rev. A* **1990**, *41*, 5414–5416.
(123) Sorouri, A.; Foulkes, W. M. C.; Hine, N. D. M. *J. Chem. Phys.* **2006**, *124*, 064105.
(124) Paier, J. *Phys. Rev. A* **2009**, *80*, 174114.
(125) Rehder, D.; Polenova, T.; Bühl, M. In *Annu. Rep. NMR Spectrosc.*; Webb, G. A., Ed. Academic Press, 2007; Vol. 64, pp. 49–114.
(126) Bühl, M.; Hamprecht, F. *J. Comput. Chem.* **1998**, *19*, 113–122.
(127) Grigoleit, S.; Bühl, M. *Chem. Eur. J.* **2004**, *10*, 5541–5552.
(128) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 8425–8433.
(129) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.
(130) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.
(131) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405–3416.
(132) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2002**, *116*, 9559.
(133) Tao, J.; Perdew, J.; Staroverov, V.; Scuseria, G. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(134) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
(135) Bühl, M.; Kabrede, H. *J. Chem. Theory Comput.* **2006**, *2*, 1282–1290.
(136) Jiménez-Hoyos, C. A.; Janesko, B. G.; Scuseria, G. E. *J. Phys. Chem. A* **2009**, *113*, 11742–11749.
(137) Sefzik, T. H.; Turco, D.; Iuliucci, R. J.; Facelli, J. C. *J. Phys. Chem. A* **2005**, *109*, 1180–1187.
(138) Schwerdtfeger, P.; Bast, R.; Gerry, M. C. L.; Jacob, C. R.; Jansen, M.; Kellö, V.; Mudring, A. V.; Sadlej, A. J.; Saue, T.; Söhnel, T.; Wagner, F. E. *J. Chem. Phys.* **2005**, *122*, 124317–9.
(139) Kellö, V.; Sadlej, A. J.; Pyykkö, P.; Sundholm, D.; Tokman, M. *Chem. Phys. Lett.* **1999**, *304*, 414–422.
(140) Cuny, J.; Messaoudi, S.; Alonzo, V.; Furet, E.; Halet, J.-F.; Le Fur, E.; Ashbrook, S. E.; Pickard, C. J.; Gautier, R.; Le Polles, L. *J. Comput. Chem.* **2008**, *29*, 2279–2287.
(141) Schwerdtfeger, P.; Pernpointner, M.; Laerdahl, J. K. *J. Chem. Phys.* **1999**, *111*, 3357–3364.
(142) Bast, R.; Schwerdtfeger, P. *J. Chem. Phys.* **2003**, *119*, 5988–5994.
(143) Thierfelder, C. *Phys. Rev. A* **2007**, *76*, 034502.
(144) Pyykkö, P. *Mol. Phys.* **2008**, *106*, 1965–1974.
(145) Bjornsson, R.; Bühl, M. *Dalton Trans.* **2010**, *39*, 5319–5324.
(146) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. *Comput. Phys. Commun.* **2010**, *181*, 1477–1489.
(147) Neese, F. *WIREs Comput. Mol. Sci.* **2011**, *2*, 73–78.
(148) Keck, K. S.; Tanjaroon, C.; Kukolich, S. G. *J. Mol. Spectrosc.* **2005**, *232*, 55–60.
(149) Daly, A. M.; Lavin, C. M.; Weidenschilling, E. S.; Holden, A. M.; Kukolich, S. G. *J. Mol. Spectrosc.* **2011**, *267*, 172–177.
(150) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
(151) Kukolich, S. G.; Sickafoose, S. M.; Flores, L. D.; Breckenridge, S. M. *J. Chem. Phys.* **1994**, *100*, 6125–6128.
(152) Høg, J.; Pedersen, T. *J. Mol. Spectrosc.* **1976**, *61*, 243–248.

(153) Kukolich, S. G.; Sickafoose, S. M. *Inorg. Chem.* **1994**, *33*, 1217–1219.
(154) Roehrig, M. A.; Wallace, D. W.; Kukolich, S. G. *J. Chem. Phys.* **1992**, *96*, 2449–2452.
(155) Kukolich, S. G.; Sickafoose, S. M. *J. Chem. Phys.* **1996**, *105*, 3466–3471.
(156) Kukolich, S. G.; Roehrig, M. A.; Haubrich, S. T.; Shea, J. A. *J. Chem. Phys.* **1991**, *94*, 191–194.
(157) Mckay, R. T.; Hubbard, J. L.; Kukolich, S. G. *J. Mol. Spectrosc.* **1995**, *172*, 378–383.
(158) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.
(159) Sickafoose, S. M.; Breckenridge, S. M.; Kukolich, S. G. *Inorg. Chem.* **2012**, *33*, 5176–5179.
(160) Javan, A.; Engelbrecht, A. *Phys. Rev.* **1954**, *96*, 649–658.
(161) Almond, M. J.; Page, E. M.; Rice, D. A.; Hagen, K.; Volden, H. V. *J. Mol. Struct.* **1994**, *319*, 223–230.
(162) Hedberg, L.; Hedberg, K.; Satija, S. K.; Swanson, B. I. *Inorg. Chem.* **1985**, *24*, 2766–2771.
(163) Almond, M. J.; Page, E. M.; Rice, D. A.; Hagen, K. *J. Organomet. Chem.* **1996**, *511*, 303–307.
(164) Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2005**, *123*, 64107.
(165) Woon, D. E.; Dunning, T. H., Jr *J. Chem. Phys.* **1995**, *103*, 4572–4585.
(166) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Constantin, L. A.; Sun, J. *Phys. Rev. Lett.* **2009**, *103*, 026403.
(167) Sinnecker, S.; Slep, L. D.; Bill, E.; Neese, F. *Inorg. Chem.* **2005**, *44*, 2245–2254.
(168) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1993**, *99*, 4597–4610.
(169) Snijders, J. G.; Sadlej, A. J. *Chem. Phys. Lett.* **1996**, *252*, 51–61.
(170) Neese, F.; Wolf, A.; Fleig, T.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2005**, *122*, 204107.
(171) Hess, B. *Phys. Rev. A* **1986**, *33*, 3742–3748.
(172) Hansen, M. R.; Madsen, G. K. H.; Jakobsen, H. J.; Skibsted, J. *J. Phys. Chem. B* **2006**, *110*, 5975–5983.
(173) Basler, W.; Lechert, H.; Paulsen, K.; Rehder, D. *J. Magn. Reson (1969)* **1981**, *45*, 170–172.
(174) Bjornsson, R.; Früchtl, H.; Bühl, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 619–627.
(175) CPMD Version 3.13.1, Copyright IBM Corp. 1990–2008, Copyright MPI für Festkörperforschung Stuttgart 1997–2001.
(176) Galy, J.; Enjalbert, R.; Jugie, G.; Strähle, J. *J. Solid State Chem.* **1983**, *47*, 143–150.
(177) Troyanov, S. I. *Russ. J. Inorg. Chem. (Transl. of Zh. Neorg. Khim.)* **2005**, *50*, 1727–1732.
(178) *R factor*; Online Dictionary of Crystallography.
(179) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2006.
(180) Kleinman, L.; Bylander, D. *Phys. Rev. Lett.* **1982**, *48*, 1425–1428.
(181) Schmidt, J.; Hoffmann, A.; Spiess, H. W.; Sebastiani, D. *J. Phys. Chem. B* **2006**, *110*, 23204–23210.
(182) Yates, J. R.; Dobbins, S. E.; Pickard, C. J.; Mauri, F.; Ghi, P. Y.; Harris, R. K. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1402–1407.
(183) Johnston, J. C.; Iuliucci, R. J.; Facelli, J. C.; Fitzgerald, G.; Mueller, K. T. *J. Chem. Phys.* **2009**, *131*, 144503.
(184) Kongsted, J.; Nielsen, C. B.; Mikkelsen, K. V.; Christiansen, O.; Ruud, K. *J.*

Chem. Phys. **2007**, *126*, 034510.
(185) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612–630.
(186) Lapina, O. B.; Mats'ko, M. A.; Mikenas, T. B.; Zakharov, V. A.; Paukshtis, E. A.; Khabibulin, D. F.; Sobolev, A. P. *Kinet. Catal.* **2001**, *42*, 553–560.
(187) Segall, M. D.; Lindan, P. J. D.; Probert, M. J.; Pickard, C. J.; Hasnip, P. J.; Clark, S. J.; Payne, M. C. *J. Phys. Condens. Matter* **2002**, *14*, 2717–2744.
(188) Pyykkö, P. *Mol. Phys.* **2001**, *99*, 1617–1629.
(189) Ferenczy, G. G.; Csonka, G. I.; Náray-Szabó, G.; Ángyán, J. G. *J. Comput. Chem.* **1998**, *19*, 38–50.
(190) Swerts, B.; Van Droogenbroeck, J.; Peeters, A.; Van Alsenoy, C. *J. Phys. Chem. A* **2002**, *106*, 4245–4250.
(191) Peeters, A.; De Maeyer, E. A. P.; Van Alsenoy, C.; Verbeeck, R. M. H. *J. Phys. Chem. B* **1997**, *101*, 3995–3998.
(192) Wen, S.; Beran, G. J. O. *J. Chem. Theory Comput.* **2011**, *7*, 3733–3742.
(193) Kimmel, A.; Muñoz Ramo, D.; Sushko, P.; Shluger, A.; Kuklja, M. *Phys. Rev. A* **2009**, *80*, 134108.
(194) Kamiński, R.; Schmøkel, M. S.; Coppens, P. *J. Phys. Chem. Lett.* **2010**, *1*, 2349–2353.
(195) Torras, J.; Bromley, S.; Bertran, O.; Illas, F. *Chem. Phys. Lett.* **2008**, *457*, 154–158.
(196) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2012**, *doi:10.1039/c2cp23949c*.
(197) Bjornsson, R.; Bühl, M. *J. Chem. Theory Comput.* **2012**, *8*, 498–508.
(198) Sokol, A. A.; French, S.; Wilson, D.; DFRL - RI, 2005, Construct, program for preparing Shell based input files for ChemShell (Written mostly by Alexey Sokol).
(199) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
(200) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.
(201) Senthilkumar, K.; Mujika, J. I.; Ranaghan, K. E.; Manby, F. R.; Mulholland, A. J.; Harvey, J. N. *J. R. Soc. Interface* **2008**, *5 Suppl 3*, S207–16.
(202) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. In *Annu. Rep. Comp. Chem.*; Wheeler, D. C. S. A. R., Ed. Elsevier, 2007; Vol. 3, pp. 177–193.
(203) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632–672.
(204) Reeve, S. W.; Burns, W. A.; Lovas, F. J.; Suenram, R. D.; Leopold, K. R. *J. Phys. Chem.* **1993**, *97*, 10630–10637.
(205) Burns, W. A.; Leopold, K. R. *J. Am. Chem. Soc.* **1993**, *115*, 11622–11623.
(206) Bühl, M.; Steinke, T.; Ragué Schleyer, von, P.; Boese, R. *Angew. Chem. Int. Ed.* **1991**, *30*, 1160–1161.
(207) Cremer, D.; Olsson, L.; Reichel, F.; Kraka, E. *Isr. J. Chem* **1993**, *33*, 369–385.
(208) Klooster, W. T.; Koetzle, T. F.; Siegbahn, P. E. M.; Richardson, T. B.; Crabtree, R. H. *J. Am. Chem. Soc.* **1999**, *121*, 6337–6343.
(209) Jiao, H.; Schleyer, P. V. R. *J. Am. Chem. Soc.* **1994**, *116*, 7429–7430.
(210) Venter, G.; Dillen, J. *J. Phys. Chem. A* **2004**, *108*, 8378–8384.
(211) Phillips, J. A.; Cramer, C. J. *J. Chem. Theory Comput.* **2005**, *1*, 827–833.
(212) Janesko, B. G. *J. Chem. Theory Comput.* **2010**, *6*, 1825–1833.
(213) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P.; Eschrig, H., Eds. Akademie Verlag: Berlin, 1991.
(214) Calhorda, M. J.; Pregosin, P. S.; Veiros, L. F. *J. Chem. Theory Comput.* **2007**, *3*,

665–670.
(215) Fernández, I.; Hermatschweiler, R.; Breher, F.; Pregosin, P. S.; Veiros, L. F.; Calhorda, M. J. *Angew. Chem. Int. Ed.* **2006**, *45*, 6386–6391.
(216) Waller, M. P.; Braun, H.; Hojdis, N.; Bühl, M. *J. Chem. Theory Comput.* **2007**, *3*, 2234–2242.
(217) Bühl, M.; Reimann, C.; Pantazis, D. A.; Bredow, T.; Neese, F. *J. Chem. Theory Comput.* **2008**, *4*, 1449–1459.
(218) Kulkarni, A. D.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 2325–2332.
(219) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
(220) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2008**, *128*, 184109.
(221) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
(222) Jacobsen, H. *J. Chem. Theory Comput.* **2011**, *7*, 3019–3025.
(223) Moellmann, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8500–8504.
(224) Cornman, C. R.; Colpas, G. J.; Hoeschele, J. D.; Kampf, J.; Pecoraro, V. L. *J. Am. Chem. Soc.* **1992**, *114*, 9925–9933.
(225) Geethalakshmi, K. R.; Waller, M. P.; Bühl, M. *Inorg. Chem.* **2007**, *46*, 11297–11307.
(226) Proctor, W.; Yu, F. *Phys. Rev.* **1951**, *81*, 20–30.
(227) Grigoleit, S.; Bühl, M. *J. Chem. Theory Comput.* **2005**, *1*, 181–193.
(228) Medek, A.; Frydman, V.; Frydman, L. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14237–14242.
(229) Dittrich, B.; Koritsanszky, T.; Volkov, A.; Mebs, S.; Luger, P. *Angew. Chem. Int. Ed.* **2007**, *46*, 2935–2938.
(230) Power, W. P.; Kirby, C. W.; Taylor, N. J. *J. Am. Chem. Soc.* **1998**, *120*, 9428–9434.
(231) Medek, A.; Frydman, L. *J. Am. Chem. Soc.* **2000**, *122*, 684–691.
(232) Jensen, J. H.; Gordon, M. S. *J. Chem. Phys.* **1998**, *108*, 4772–4782.
(233) Sebastiani, D.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 2807–2815.
(234) Fox, S. J.; Pittock, C.; Fox, T.; Tautermann, C. S.; Malcolm, N.; Skylaris, C.-K. *J. Chem. Phys.* **2011**, *135*, 224107.
(235) Fradelos, G.; Wesołowski, T. A. *J. Phys. Chem. A* **2011**, *115*, 10018–10026.
(236) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534–10547.
(237) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 3330–3342.
(238) Wang, B.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10556–10564.
(239) Paricaud, P.; Předota, M.; Chialvo, A. A.; Cummings, P. T. *J. Chem. Phys.* **2005**, *122*, 244511.
(240) Biswas, P. K.; Gogonea, V. *J. Chem. Phys.* **2005**, *123*, 164114.
(241) Jacob, C. R.; Visscher, L. *J. Chem. Phys.* **2006**, *125*, 194104.
(242) Hratchian, H. P.; Krukau, A. V.; Parandekar, P. V.; Frisch, M. J.; Raghavachari, K. *J. Chem. Phys.* **2011**, *135*, 014105.
(243) Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valentine, J. S. *Biological inorganic chemistry*; University Science Books Mill Valley, 2007.
(244) Rehder, D. *Angew. Chem. Int. Ed.* **1991**, *30*, 148–167.
(245) Rehder, D. *Biometals* **1992**, *5*, 3–12.
(246) Rehder, D. *Inorg. Chem. Commun.* **2003**, *6*, 604–617.
(247) Crans, D. C.; Smee, J. J.; Gaidamauskas, E.; Yang, L. *Chem. Rev.* **2004**, *104*, 849–902.
(248) Rehder, D. *Bioinorganic vanadium chemistry*; Wiley, 2008.
(249) Thompson, K. H.; Orvig, C. *J. Inorg. Biochem.* **2006**, *100*, 1925–1935.
(250) Thompson, K. H.; Lichter, J.; LeBel, C.; Scaife, M. C.; McNeill, J. H.; Orvig, C. *J.*

Inorg. Biochem. **2009**, *103*, 554–558.
(251) Blasiak, L. C.; Drennan, C. L. *Acc. Chem. Res.* **2009**, *42*, 147–155.
(252) Butler, A. *Science* **1998**, *281*, 207–210.
(253) Butler, A.; Carrano, C. J. *Coord. Chem. Rev.* **1991**, *109*, 61–105.
(254) van de Velde, F.; Arends, I. W. C. E.; Sheldon, R. A. *J. Inorg. Biochem.* **2000**, *80*, 81–89.
(255) Ligtenbarg, A. *Coord. Chem. Rev.* **2003**, *237*, 89–101.
(256) Messerschmidt, A.; Wever, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 392–396.
(257) Renirie, R.; Charnock, J. M.; Garner, C. D.; Wever, R. *J. Inorg. Biochem.* **2010**, *104*, 657–664.
(258) Conte, V.; Di Furia, F.; Moro, S.; Rabbolini, S. *J. Mol. Catal. A: Chem.* **1996**, *113*, 175–184.
(259) Conte, V.; Di Furia, F.; Moro, S. *J. Mol. Catal. A: Chem.* **1997**, *120*, 93–99.
(260) Conte, V.; Bortolini, O.; Carraro, M.; Moro, S. *J. Inorg. Biochem.* **2000**, *80*, 41–49.
(261) Bortolini, O.; Carraro, M.; Conte, V.; Moro, S. *Eur. J. Inorg. Chem.* **2003**, *2003*, 42–46.
(262) Pecoraro, V. L.; Kravitz, J. Y. *Pure Appl. Chem.* **2005**, *77*, 1595–1606.
(263) Zampella, G.; Kravitz, J. Y.; Webster, C. E.; Fantucci, P.; Hall, M. B.; Carlson, H. A.; Pecoraro, V. L.; De Luca, L. *Inorg. Chem.* **2004**, *43*, 4127–4136.
(264) Kravitz, J. Y.; Pecoraro, V. L.; Carlson, H. A. *J. Chem. Theory Comput.* **2005**, *1*, 1265–1274.
(265) Raugei, S.; Carloni, P. *J. Phys. Chem. B* **2006**, *110*, 3747–3758.
(266) Waller, M. P.; Bühl, M.; Geethalakshmi, K. R.; Wang, D.; Thiel, W. *Chem. Eur. J.* **2007**, *13*, 4723–4732.
(267) Waller, M. P.; Geethalakshmi, K. R.; Bühl, M. *J. Phys. Chem. B* **2008**, *112*, 5813–5823.
(268) Geethalakshmi, K. R.; Waller, M. P.; Thiel, W.; Bühl, M. *J. Phys. Chem. B* **2009**, *113*, 4456–4465.
(269) Zhang, Y.; Gascón, J. A. *J. Inorg. Biochem.* **2008**, *102*, 1684–1690.
(270) Macedo-Ribeiro, S.; Hemrika, W.; Renirie, R.; Wever, R.; Messerschmidt, A. *J. Biol. Inorg. Chem.* **1999**, *4*, 209–219.
(271) de Boer, E.; Wever, R. *J. Biol. Chem.* **1988**, *263*, 12326–12332.
(272) Van Schijndel, J. W.; Barnett, P.; Roelse, J.; Vollenbroek, E. G.; Wever, R. *Eur. J. Biochem.* **1994**, *225*, 151–157.
(273) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
(274) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(275) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.
(276) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
(277) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14*, 136–141.
(278) *IUPAC Compendium of Chemical Terminology*; Gold Book; 2nd ed. IUPAC: Research Triagle Park, NC, 2009.
(279) Froestl, W. *Future Med. Chem.* **2011**, *3*, 163–175.

(280) O'Hagan, D. *Future Med. Chem.* **2011**, *3*, 189–195.
(281) Deniau, G.; Slawin, A. M. Z.; Lebl, T.; Chorki, F.; Issberner, J. P.; van Mourik, T.; Heygate, J. M.; Lambert, J. J.; Etherington, L.-A.; Sillar, K. T.; O'Hagan, D. *ChemBioChem* **2007**, *8*, 2265–2274.
(282) Mazzanti, A.; Casarini, D. *WIREs Comput. Mol. Sci.* **2011**, *doi:10.1002/wcms.96*.
(283) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351–361.
(284) Price, W. D.; Jockusch, R. A.; Williams, E. R. *J. Am. Chem. Soc.* **1997**, *119*, 11988–11989.
(285) Xu, S.; Nilles, J. M.; Bowen Kit H, J. *J. Chem. Phys.* **2003**, *119*, 10696–10701.
(286) Crittenden, D. L.; Chebib, M.; Jordan, M. J. T. *J. Phys. Chem. A* **2004**, *108*, 203–211.
(287) Tse, Y. C.; Newton, M. D.; Vishveshwara, S.; Pople, J. A. *J. Am. Chem. Soc.* **1978**, *100*, 4329–4331.
(288) Tortonda, F. R.; Pascual-Ahuir, J.-L.; Silla, E.; Tuñón, I.; Ramirez, F. J. *J. Chem. Phys.* **1998**, *109*, 592–603.
(289) Leung, K.; Rempe, S. B. *J. Chem. Phys.* **2005**, *122*, 184506.
(290) Lu, Z.; Zhang, Y. *J. Chem. Theory Comput.* **2008**, *4*, 1237–1248.
(291) Crittenden, D. L.; Chebib, M.; Jordan, M. J. T. *J. Phys. Chem. A* **2005**, *109*, 4195–4201.
(292) Crittenden, D. L.; Kumar, R. J.; Hanrahan, J. R.; Chebib, M.; Jordan, M. J. T. *J. Phys. Chem. A* **2005**, *109*, 8398–8409.
(293) Cao, J.; Bjornsson, R.; Bühl, M.; Thiel, W.; van Mourik, T. *Chem. Eur. J.* **2012**, *18*, 184–195.
(294) Thiel, W. Program MNDO.
(295) Schiferl, S. K.; Wallace, D. C. *J. Chem. Phys.* **1985**, *83*, 5203–5209.
(296) Helgaker, T.; Jaszuński, M.; Ruud, K.; Górska, A. *Theor. Chem. Acc.* **1998**, *99*, 175–182.
(297) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.
(298) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283–290.
(299) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–124.
(300) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105.
(301) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173–1213.
(302) Štrajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333–13343.
(303) Plotnikov, N. V.; Kamerlin, S. C. L.; Warshel, A. *J. Phys. Chem. B* **2011**, *115*, 7950–7962.
(304) Kamerlin, S. C. L.; Warshel, A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10401–10411.
(305) Woods, C. J.; Manby, F. R.; Mulholland, A. J. *J. Chem. Phys.* **2008**, *128*, 014109.
(306) Retegan, M.; Martins-Costa, M.; Ruiz-López, M. F. *J. Chem. Phys.* **2010**, *133*, 064103.
(307) Beierlein, F. R.; Michel, J.; Essex, J. W. *J. Phys. Chem. B* **2011**, *115*, 4911–4926.
(308) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
(309) Ho, J.; Klamt, A.; Coote, M. L. *J. Phys. Chem. A* **2010**, *114*, 13442–13444.
(310) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2011**, *115*, 14556–14562.
(311) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *ChemPhysChem* **2009**, *10*,

1125–1134.
(312) Blanco, S.; López, J. C.; Mata, S.; Alonso, J. L. *Angew. Chem. Int. Ed.* **2010**, *49*, 9187–9192.
(313) Kästner, J. *WIREs Comput. Mol. Sci.* **2011**, *1*, 932–942.
(314) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2008**, *10*, 297–302.
(315) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2007**, *3*, 1499–1509.
(316) Hofer, T. S.; Pribil, A. B.; Randolf, B. R.; Rode, B. M. *J. Am. Chem. Soc.* **2005**, *127*, 14231–14238.
(317) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231–2241.
(318) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *J. Chem. Theory Comput.* **2009**, *5*, 2212–2221.
(319) Bernstein, N.; Várnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csányi, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 646–656.
(320) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. *Chem. Phys. Lett.* **2012**, *524*, 56–61.
(321) Korth, M.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 2929–2936.
(322) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
(323) Garrett, R. H.; Grisham, C. M. *Biochemistry*; 4th ed. Brooks/Cole, 2010; pp. 119–120, 993.
(324) Kang, H. J.; Baker, E. N. *Trends Biochem. Sci.* **2010**, 1–9.
(325) Wikoff, W. R.; Liljas, L.; Duda, R. L.; Tsuruta, H.; Hendrix, R. W.; Johnson, J. E. *Science* **2000**, *289*, 2129–2133.
(326) Kang, H. J.; Coulibaly, F.; Clow, F.; Proft, T.; Baker, E. N. *Science* **2007**, *318*, 1625–1628.
(327) Kang, H. J.; Baker, E. N. *J. Biol. Chem.* **2009**, *284*, 20729–20737.
(328) Oke, M.; Carter, L. G.; Johnson, K. A.; Liu, H.; McMahon, S. A.; Yan, X.; Kerou, M.; Weikart, N. D.; Kadi, N.; Sheikh, M. A.; Schmelz, S.; Dorward, M.; Zawadzki, M.; Cozens, C.; Falconer, H.; Powers, H.; Overton, I. M.; van Niekerk, C. A. J.; Peng, X.; Patel, P.; Garrett, R. A.; Prangishvili, D.; Botting, C. H.; Coote, P. J.; Dryden, D. T. F.; Barton, G. J.; Schwarz-Linek, U.; Challis, G. L.; Taylor, G. L.; White, M. F.; Naismith, J. H. *J. Struct. Funct. Genomics* **2010**, *11*, 167–180.
(329) Hagan, R. M.; Bjornsson, R.; McMahon, S. A.; Schomburg, B.; Braithwaite, V.; Bühl, M.; Naismith, J. H.; Schwarz-Linek, U. *Angew. Chem. Int. Ed.* **2010**, *49*, 8421–8425.
(330) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins Struct. Funct. Bioinf.* **2005**, *61*, 704–721.
(331) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. *J. Phys. Chem. A* **2009**, *113*, 11856–11865.
(332) Jónsson, H.; Mills, G.; Jacobsen, K. W. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J.; Ciccotti, G.; Coker, D. F., Eds. World Scientific, 1998; p. 385.
(333) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978–9985.
(334) Goumans, T. P. M.; Catlow, C. R. A.; Brown, W. A.; Kästner, J.; Sherwood, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 5431–5436.
(335) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111*, 7010–7022.
(336) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; Jónsson, H. *J. Chem. Phys.* **2004**, *121*, 9776–9792.
(337) Heyden, A.; Bell, A. T.; Keil, F. J. *J. Chem. Phys.* **2005**, *123*, 224101.

(338) Kästner, J.; Sherwood, P. *J. Chem. Phys.* **2008**, *128*, 014106.
(339) Thiel, W. In *Multiscale Simulation Methods in Molecular Sciences*; Grotendorst, J.; Attig, N.; Blügel, S.; Marx, D., Eds. John von Neuman Institute for Computing: Jülich, 2009; Vol. 42, pp. 203–214.
(340) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; illustrated edition. University Science: USA, 2005; pp. 511–545.
(341) Pointon, J. A.; Smith, W. D.; Saalbach, G.; Crow, A.; Kehoe, M. A.; Banfield, M. J. *J. Biol. Chem.* **2010**, *285*, 33858–33866.
(342) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. *J. Am. Chem. Soc.* **2011**, *133*, 478–485.
(343) Peng, C.; Schlegel, H. B. *Isr. J. Chem* **1993**, *33*, 449–454.
(344) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452–461.
(345) Senn, H. M.; Kästner, J.; Breidung, J.; Thiel, W. *Can. J. Chem.* **2009**, *87*, 1322–1337.
(346) Pattabiraman, V. R.; Bode, J. W. *Nature* **2011**, *480*, 471–479.
(347) Schneider, H.-J. *Angew. Chem. Int. Ed.* **2009**, *48*, 3924–3977.
(348) Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7897–7901.
(349) Schmeing, T. M.; Huang, K. S.; Kitchen, D. E.; Strobel, S. A.; Steitz, T. A. *Mol. Cell* **2005**, *20*, 437–448.
(350) Brunelle, J. L.; Shaw, J. J.; Youngman, E. M.; Green, R. *RNA* **2008**, *14*, 1526–1531.
(351) Kästner, J.; Sherwood, P. *Mol. Phys.* **2010**, *108*, 293–306.
(352) Wallin, G.; Aqvist, J. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1888–1893.
(353) Huang, Y.; Sprinzl, M. *Angew. Chem. Int. Ed.* **2011**, *50*, 7287–7289.
(354) Hiller, D. A.; Singh, V.; Zhong, M.; Strobel, S. A. *Nature* **2011**, *476*, 236–239.
(355) Pech, M.; Nierhaus, K. H. *ChemBioChem* **2012**, *13*, 189–192.

Appendix

Appendix 1.1:

The QM/MM protocol in Chapter 3.5 relied on the use of several Chemshell[89], TCl and bash scripts. These are included below:

Chemshell example inputfile for $VOCl_3$:

```
##########################################
# COORDINATES INPUT
##########################################

global chemsh_default_connectivity_toler
global chemsh_default_connectivity_scale
#set chemsh_default_connectivity_toler  X
#set chemsh_default_connectivity_scale  Y

c_create coords=cell.c {
cell_constants angstrom
5.02300 9.21600 11.27800 90.00000 90.00000 90.000
space_group
1
coordinates
V    0.141410   0.750000   0.954780
V    0.641410   0.750000   0.545220
V    0.858590   0.250000   0.045220
V    0.358590   0.250000   0.454780
O    0.829100   0.750000   0.950500
O    0.329100   0.750000   0.549500
O    0.170900   0.250000   0.049500
O    0.670900   0.250000   0.450500
Cl   0.255800   0.750000   0.136480
Cl   0.755800   0.750000   0.363520
Cl   0.744200   0.250000   0.863520
Cl   0.244200   0.250000   0.636480
Cl   0.273800   0.558160   0.867110
Cl   0.773800   0.941840   0.632890
Cl   0.273800   0.941840   0.867110
Cl   0.773800   0.558160   0.632890
Cl   0.726200   0.441840   0.132890
Cl   0.226200   0.058160   0.367110
Cl   0.726200   0.058160   0.132890
Cl   0.226200   0.441840   0.367110
}

##################################################
# CALCULATION SETTINGS
##################################################

source sc-qmmm.tcl

## QM calculation settings
set moleculecharge 0
set qmatoms {1 2 3 4 5 }
exec touch ./empty.ff


sc-qmmm cell=cell.c   \
        atomchargesupdate=./chargeupdate-npa-g09.bash \
        cellchargesupdate=./chargecell.bash \
        radiuscluster=80 \
        radius_active=50 \
        originatom=1 \
        nsp=10 \
        nopt=10 \
        rmscon=0.00001 \
        chargelevel= [ list theory=hybrid : [ list coupling=shift  \
                qm_region= $qmatoms \
                qm_theory= gaussian : [ list  \
                        exe= g09run \
                        nproc= 8  \
```

```
                    scfconv= 6 \
                    basis=6-31G* \
                    charge=$moleculecharge \
                    hamiltonian= bp86 \
                    optstr= "int=ultrafine" \
                    atomcharge= pop=npa \
                    g98_mem= 1GB ] \
        mm_theory=dl_poly :  [ list  \
                     mxexcl= 254 \
                     mxlist= 20000 \
                     mm_defs= empty.ff  ]]] \
    optlevel= [ list tolerance=0.0045  \
        active_atoms= $qmatoms maxcycle=700 \
        theory=hybrid : [ list coupling=shift  \
        qm_region= $qmatoms \
        qm_theory= orca : [ list  \
                   executable=/home/rb465/orca_280-update2/orca \
                    nproc=8 \
                    hamiltonian= dft \
                    functional=bp86 \
                    use_ri=yes \
                    basis=def2svp \
                    auxbasis=def2_SVP_J \
                    gridsize=4 \
                    finalgridsize=5 ] \
        mm_theory=dl_poly :  [ list  \
                   mxexcl= 254 \
                   mxlist= 20000 \
                   mm_defs= vocl3.ff  ]]]
```

sc-qmmm.tcl:

```
proc sc-qmmm { args } {

  if { [ parsearg sc-qmmm {cell chargelevel optlevel \
                    atomchargesupdate cellchargesupdate  radiuscluster \
                    radius_active originatom nsp nopt rmscon } $args ] } {
     chemerr "error in sc-qmmmm arguments"
  }


#####################################################################
# CLUSTER CUTTING
#####################################################################
cluster_cut coords=cell.c radius_cluster=$radiuscluster cluster=cluster.c crystal_type=molecular origin_atom=$originatom

exec cp cluster.c cluster-initial.c

#####################################################################
# SINGLE-POINT LOOP
#####################################################################

for { set isp 1 } { $isp <= $nsp } { incr isp 1 } {

set dsp [expr $isp-1]
# Single-point calculation to get charges

energy coords=cluster.c $chargelevel
#
exec cp cluster.c cluster-afterenergy$isp.c

# Updating point charges in fragment file (cluster.c)
exec $atomchargesupdate cluster
exec cp cluster.c cluster-afterupdate$isp.c
exec mv fieldfile fieldfile$isp

# Assigning the atom charges to fractional_atom charges by a DL_POLY energy calculation
energy coords=cluster.c theory=dl_poly : [ list list_option=full mm_defs=empty.ff save_charges=yes ]
exec cp cluster.c cluster-afterenergydlpoly$isp.c


# Copying fractional atom_charges into cell.c
exec cp cell.c cell.c.$isp.preupdate
exec $cellchargesupdate
exec cp      cell.c cell.c.$isp.afterupdate
```

```
# Fit background charges
cluster_fit_bqs bq_symbol=Bq  radius_active=$radius_active crystal=$cell cluster=cluster.c cluster_origin=cluster_origin
exec cp cluster.c cluster-afterfit$isp.c


# Outputting new cluster after each cycle
exec cp cluster.c cluster-sp$isp.c

puts "Single-point cycle nr. $isp done !"
if { ($isp > 1) } { set rmscharge [get_rms_difference left=fieldfile$dsp right=fieldfile$isp ]
puts stdout "RMSD on charges:"
puts stdout $rmscharge }
if { ($isp > 1)  &&  ($rmscharge < $rmscon) } {
puts stdout "RMSD on charges, less            than $rmscon, exiting Single-point loop..."
break}
if { ($isp == $nsp)  } {
puts stdout "Maximum number of cycles reached, charges not converged."
error "Exiting due to max cycles reached" }

}
#####################################################################
# OPTIMISATION LOOP
#####################################################################

exec cp cluster.c cluster-opt0-fit.c

for { set iopt 0 } { $iopt <= $nopt } { incr iopt 1 } {

set nextopt [ expr $iopt + 1 ]
set dopt [expr $isp-1]

dl-find coords=cluster-opt$iopt-fit.c $optlevel result=cluster-opt$nextopt.c


# Single-point calculation to get charges
energy $chargelevel coords=cluster-opt$nextopt.c

# Updating point charges in fragment file (cluster.c)
exec $atomchargesupdate cluster-opt$nextopt

# Assigning the atom charges to fractional_atom charges by a DL_POLY energy calculation
energy coords=cluster-opt$nextopt.c theory=dl_poly : [ list list_option=full mm_defs=empty.ff save_charges=yes ]

# Copying fractional atom_charges into cell.c
exec $cellchargesupdate

# Fit background charges
cluster_fit_bqs bq_symbol=Bq  radius_active=$radius_active crystal=cell.c cluster=cluster-opt$nextopt-fit.c
cluster_origin=cluster_origin


puts "Optimisation cycle nr. $iopt done !"
if { ($iopt > 1) } { set rmscharge [get_rms_difference left=fieldfile$dopt right=fieldfile$iopt ]
puts stdout "RMSD on charges:"
puts stdout $rmscharge }
if { ($iopt > 1)  &&  ($rmscharge < $rmscon) } {
puts stdout "RMSD on charges, less than $rmscon, exiting Optimisation loop."
puts stdout "Self-consistent QM/MM optimisation complete!"
exec cp cluster-opt$iopt-fit.c result.c
break}
if { ($iopt == $nopt)  } {
puts stdout "Maximum number of optimisation cycles reached, charges not converged."
error "Exiting due to optimisation max cycles reached" }


}

exec rm [glob fieldfile*]
}
```

chargecell.bash:

```
#!/bin/bash


# Taking cell charges from cluster.c
sed  -n '/fractional_atom_charges/,/xtal_map  records/p' cluster.c > cellcharges-pre
```

```
sed '1d' cellcharges-pre | sed  '$d' > cellcharges

numcellcharges=`wc -l cellcharges | awk '{print $1}'`

prevcharges=`grep atom_charges cell.c`
# Deleting previous charges

if [ "$prevcharges" == "" ]; then
echo "No previous charges"
else
sed -i '/block = atom_charges/,/block = xtal_map/{//p;d;}' cell.c
sed -i '/block = atom_charges/d' cell.c
fi

# Adding charges to cell.c
sed -i "/xtal_map/ i\
block = atom_charges records = $numcellcharges" cell.c
sed -i '/block = atom_charges records/a  something' cell.c
sed -i '/something/ {r ./cellcharges
d}' cell.c

sed -i "/xtal_map/ i\
block = fractional_atom_charges records = $numcellcharges" cell.c
sed -i '/block = fractional_atom_charges records/a  something' cell.c
sed -i '/something/ {r ./cellcharges
d}' cell.c


 rm cellcharges-pre cellcharges
```

chargeupdate-npa-g09.bash:

```
#!/bin/bash

clusterfile=$1

# Gaussian 09 NPA specific settings
natoms=`grep -m 1 NAtoms gaussian.log | awk '{print $2}'`
natoms2=`expr $natoms + 2`
grep -A$natoms2 "Atom  No    Charge" gaussian.log >zurg
sed -n -i '/---------------------/,/==================/p' zurg
sed -i '1d' zurg
sed -i '$d' zurg
awk '{print $3}' zurg> zorg1
atommol=`wc -l zorg1 | awk '{print $1}'`

# Creating field file
echo "block = field records = 0" >fieldfile
echo "block = field_grid records = $natoms elements = 3" >>fieldfile
sed  -n -e '/coordinates/,/charges/ { /coordinates/b; /records/b; p }' hybrid.gaussian.coords >fieldcoords
awk '{print $2"   "$3"   "$4}' fieldcoords >>fieldfile
echo "block = field_data records = $natoms elements = 1" >>fieldfile
cat zorg1 >>fieldfile

# Settings to deal with Bq charges
bq=`grep -c Bq $clusterfile.c`
atomwithbq=`grep 'block = atom_charges records =' $clusterfile.c | awk '{print $6}'`
atomfull=`expr $atomwithbq - $bq`

#Deleting xtal Bq data
sed  -n '/block = xtal_map/,/block = connectivity/p' $clusterfile.c > xtalstuff
sed -i '/-1/d' xtalstuff
sed -i '$d' xtalstuff
sed  -i '/block = xtal_map/,/block = connectivity/{//p;d;}' $clusterfile.c

# Adding marker to original file
sed -i '/block = xtal_map/a  xtal' $clusterfile.c
# Adding xtal data back
sed -i '/xtal/ {r ./xtalstuff
d}' $clusterfile.c


# Deleting all atom_charges
sed -i '/atom_charges records/,/fractional_atom_charges/{//p;d;}' $clusterfile.c

# Taking all coordinates
```

```
sed  -n '/coordinates records/,/fractional_coordinates records/p' $clusterfile.c >atomcoordinates-pre
sed '1d' atomcoordinates-pre | sed  '$d' > atomcoordinates

# Deleting all atom coordinates
sed -i '/block = coordinates records =/,/block = fractional_coordinates records/{//p;d;}' $clusterfile.c

# Deleting Bq charges from fragment file
sed -i '/Bq/d' atomcoordinates

cat atomcoordinates >fullcoordinates


# Adding marker to original file
sed -i '/block = coordinates records =/a  something' $clusterfile.c

sed -i '/something/ {r ./fullcoordinates
d}' $clusterfile.c


numcomplexatoms=$atomfull
nummol=`expr $numcomplexatoms / $atommol`


# # Making full charge-list, Gaussian 09
counter=1
cat /dev/null > output
while (( $counter <= $nummol )); do
      cat zorg1  >> output
      let counter=$counter+1
done
cat output >allcharges

# Adding all charges to file.
sed -i '/block = atom_charges records/a  something' $clusterfile.c
sed -i '/something/ {r ./allcharges
d}' $clusterfile.c

# Changing numbers in $clusterfile.c due to deletion of Bq charges (if present)
if [ "$bq" == "" ]; then
echo "No Bq charges present in first loop"
else
sed -i "s/block = coordinates records = $atomwithbq/block = coordinates records = $atomfull/g" $clusterfile.c
sed -i "s/block = atom_charges records = $atomwithbq/block = atom_charges records = $atomfull/g" $clusterfile.c
sed -i "s/block = xtal_map  records = $atomwithbq/block = xtal_map  records = $atomfull/g" $clusterfile.c
fi
cp $clusterfile.c $clusterfile-fit.c
rm fieldcoords zurg output  atomcoordinates-pre atomcoordinates fullcoordinates allcharges
```