

Low-Scaling Local and Fragment Self-Consistent Field Potentials in Molecular Systems

Dissertation
for the award of the degree
"Doctor rerum naturalium" (Dr.rer.nat.)
of the Georg-August-Universität Göttingen

within the doctoral program Chemistry
of the Georg-August University School of Science (GAUSS)

submitted by
Martin Werner
from Hofgeismar

Göttingen, 2017

Thesis Committee

Prof. Dr. Ricardo Mata
Institute for Physical Chemistry,
Georg-August University Göttingen

Prof. Dr. Martin Suhm
Institute for Physical Chemistry,
Georg-August University Göttingen

Members of the Examination Board

Reviewer:

Prof. Dr. Ricardo Mata
Institute for Physical Chemistry,
Georg-August University Göttingen

Second Reviewer:

Prof. Dr. Martin Suhm
Institute for Physical Chemistry,
Georg-August University Göttingen

Further members of the Examination Board

Prof. Dr. Jörg Behler
Institute for Physical Chemistry,
Georg-August University Göttingen

apl. Prof. Dr. Bert de Groot
Computational Biomolecular Dynamics Group,
Max Planck Institute for Biophysical Chemistry Göttingen

Prof. Dr. Inke Siewert
Institute for Inorganic Chemistry,
Georg-August University Göttingen

Prof. Dr. Ivo Feussner
Department of Plant Biochemistry,
Albrecht-von-Haller-Institute for Plant Sciences Göttingen

Date of the oral examination:

24.11.2017

It is nice to know that the computer
understands the problem, but I would
like to understand it too.

Eugene Paul Wigner
(Nobel laureate in Physics 1963)

Acknowledgements

First of all, I would like to thank Prof. Dr. Ricardo Mata for the supervision of this PhD project. He supported my research carrier since the beginning of my bachelor studies and was always open to all of my questions arising during these years. Moreover, he introduced me to the interesting research field of Theoretical Chemistry in biomolecular applications, which eventually resulted in the subject of the thesis at hand. It is a pleasure to have Prof. Dr. Martin Suhm as the second reviewer for this thesis and I would like to express my gratitude to him.

The major part of my work would have been impossible without discussions and collaborations with several colleagues. In this context I would like to thank Prof. Dr. Hans-Joachim Werner for sharing the LDF project and for the opportunity to spend a month in his group in Stuttgart to work on the implementation of the open-shell LDF procedure in the MOLPRO program. In line with this, I am also indebted to Dr. Christoph Köppl, who guided me through the difficult parts of the source code with useful hints and suggestions. I would like to express my thanks to Prof. Dr. Ivo Feussner and Dr. Julia Newie for the fruitful collaboration on the enzyme selectivity study of CspLOX2. The broad discussions in the course of this project on experimental and theoretical results led to a mutual benefit and are the key element to the success of any scientific work.

Furthermore, I especially wish to express my gratitude to all current and former members of the Computational Chemistry and Biochemistry group in Göttingen for the useful discussions and remarks, the nice working atmosphere and the entertaining group activities. I appreciate the contribution of everyone, who helped me by proof-reading parts of this thesis.

I would like to thank my whole family for every kind of help that I experienced during the time of my PhD. My special thanks go to my brother Frank, who carefully proof-read this entire thesis and helped with useful comments.

Last but not least, I want to thank my wife Anna for always supporting me in every possible way. I am thankful for her patience and her appreciation of the situation, when my time was limited by the work included in this thesis.

Abstract

Due to the steep scaling of computational cost with increasing system size, approximations are required in order to make quantum mechanical calculations in large molecular systems amenable. Particular progress has been made in the field of single-reference local correlation methods allowing the post-Hartree-Fock part of the calculation to consume only a minor amount of the entire computational time. This turns the spotlight to the calculation of the reference itself, the Hartree-Fock wave function. Viable approximations are needed to accelerate the evaluation of the self-consistent field solution but at the same time conserve a converged and accurate molecular description. Two different approaches to address this issue are presented within the course of this thesis.

In the first approach a cap-free fragmentation procedure splitting up the molecular system into smaller entities is employed. One-body calculations are carried out in the full Fock potentials of the remaining structure allowing for the convergence to the Hartree-Fock limit with repetitive iterations over the generated fragments. The inclusion of full Fock potentials of the environment also provides an exact embedding procedure empowering the critical analysis of potentials applied to QM/MM structures in condensed phase.

The second approach directly addresses the most demanding step to the evaluation of the Hartree-Fock reference, the computation of the exchange integrals. Local density fitting approximations are applied to exploit the physics of exchange contributions and result in a linear-scaling algorithm for the computation of Hartree-Fock exchange. This enables the execution of accelerated self-consistent field procedures affecting the Hartree-Fock approach and due to the admixture of exact exchange also hybrid density functional methods. Of particular interest is the application of local density fitting approximations to the investigation of open-shell systems. For restricted open-shell and unrestricted formulations of Hartree-Fock and hybrid density functional theory the limitations and speed up gains of the introduced approximations are analyzed in radical benchmark structures.

In the selectivity analysis of a cyanobacterial lipoxxygenase enzyme the local density fitting approach is applied to large open-shell systems. The quantum mechanical treatment of the active site rules out electronic structure effects as a steering factor for the selective product formation. Instead, steric effects move into the focus of investigation and are analyzed through the use of a newly developed shielding model. With the criteria of a selective precursory state formation this model is the first one to quantitatively reproduce experimental product distributions of several lipoxxygenase enzyme structures.

Contents

Acknowledgements	I
Abstract	III
1 Introduction	1
2 Theoretical Background	5
2.1 Quantum Mechanics	5
2.1.1 Hartree-Fock Theory	6
2.1.2 Orbital Localization	13
2.1.3 Density Functional Theory	14
2.1.4 Density Fitting	17
2.2 Fragmentation Techniques	20
2.2.1 Molecular Fractionation with Conjugate Caps	21
2.2.2 Fragment Molecular Orbital Method	23
2.3 Molecular Mechanics	24
2.3.1 Force Fields	25
2.3.2 Molecular Dynamics	26
3 Fock Potentials in Embedded Calculations	29
3.1 One-electron Operators	29
3.2 FJK	31
3.2.1 Method	31
3.2.2 Convergence Analysis	34
3.2.3 Conformational Energy	37
3.2.4 Mixed Basis Approach	42
3.3 Embedding Analysis	46
3.3.1 Procedure	47
3.3.2 Monomer Density	48
3.3.3 Dispersion energy	53
3.4 Conclusion	55
4 Local Density Fitting Approximations	59
4.1 Method	60
4.2 SCF Approximations	62
4.3 Benchmark Calculations	64
4.3.1 Delocalized Systems	64
4.3.2 RAFT Polymerization	67
4.4 Conclusion	72
5 Theoretical Study of CspLOX2 Selectivity	75
5.1 Lipoxygenases	76

5.2	Computational Details	78
5.3	Selectivity Analysis	80
5.3.1	Spin Density	80
5.3.2	Activation Barriers	84
5.3.3	Substrate Conformation	86
5.3.4	Steric Shielding	88
5.4	Conclusion	92
6	Summary	95
	Appendix	101
A.1	FJK Benchmark Structures	101
A.2	Embedding Analysis Structures	104
A.3	CspLOX2 Model System Structures	110
A.4	Code Segment of the Steric Shielding Analysis	113
	Bibliography	115

1 Introduction

Computational Chemistry is a rapidly evolving research field with a broad spectrum of application. With the possibility to describe virtually every structure of chemical interest by theoretical models, several orders of magnitude in size and time scales are covered. The applications to systems of different size reach from small molecules in outer space up to large macromolecular compounds within biological cells, while the time scales extend from almost instantaneous electron excitation to comparably slow molecular motions. Due to the impressive developments of computational power and technology in the past decades, one might get the impression that it is only a matter of time until every chemical phenomenon will be accessible to accurate theoretical calculations. In fact, all equations for a quantum mechanical treatment of the system under consideration are known and could be in principle readily put to computation. However, such methods exhibit a steep exponential scaling with an increase of the system size, making them applicable in reasonable computational timings for small and medium-sized models only. On the rough estimation of Moore's law, according to which the CPU power doubles every two years, the high-level quantum mechanical solutions of structures beyond a very limited number of atoms will not be feasible over the course of our lifetimes.

Regarding the abovementioned issues, the fast and accurate description of large molecular systems with quantum mechanical methods is still today the holy grail of Computational Chemistry. Due to the high computational cost of such calculations, approximations are needed to break the wall of exponential scaling. The aim is to obtain quantum mechanical methods which are linear or at least low-scaling in terms of their computational timings with an increasing molecular size. Significant efforts have been made in the last decades towards low cost single-reference methods, based upon Hartree-Fock theory.^[1-4] Especially in the field of local correlation methods the recent developments have been actually so significant, that the evaluation of the Hartree-Fock wave function may act as a computational bottleneck.^[5;6] This calls for the introduction of approximations to Hartree-Fock calculations which significantly speed up the procedure without touching the accurate reference character for subsequent correlation treatments. With this purpose in mind, different approaches to obtain an accelerated self-consistent field solution are investigated.

A direct attempt is provided by fragmentation techniques which divide the total system into smaller entities and thus transfer the demanding calculation to several subsets of heavily decreased computational cost.^[7;8] With the inclusion of the effect of the remaining structure by exact potentials for the treatment of each entity, the aim is to restore the complete Hartree-Fock solution solely from simple fragment calculations. Particular effort is put towards a favorable scaling of the procedure compared to the full molecular

Hartree-Fock description. Furthermore, the employed inclusion of the effect of surrounding monomers provides an accurate potential and can be used as a reference for the analysis of other established embedding techniques. The comparison to point charge models commonly applied in QM/MM approaches of condensed matter simulations yields insights into the impact of embedding potentials on the calculated molecular properties.

Another approach towards a low-scaling self-consistent field procedure concerns the ease of computation of the most demanding part within Hartree-Fock theory, the evaluation of exchange. Due to viable approximations employing local orbital spaces, the rapid decay of exchange contributions can be exploited. This local density fitting approach enables the formulation of a linear scaling algorithm for the computation of the exchange matrix and has been recently implemented for closed-shell Hartree-Fock calculations.^[9] The extension of the approach to open-shell calculations in restricted and unrestricted formulations of Hartree-Fock theory is part of this work. Due to the admixture of exact Hartree-Fock exchange, the employment of the approach also enables potential speed ups in hybrid density functional applications. With a strict control of the errors introduced by the local density fitting approximations, the resulting procedures eventually facilitate an accurate description of large biomolecular structures with reasonable computational timings.

The cyanobacterial lipoxygenase CspLOX2 provides a system of high biochemical complexity, suitable for the application of local and fragment potentials. Lipoxygenases are a class of enzymes which play a key role in lipid-mediator biosynthesis by catalyzing the dioxygenation reaction of polyunsaturated fatty acids.^[10;11] CspLOX2 exhibits an extraordinary selectivity in terms of the dioxygenation products of its fatty acid substrate. Different single amino acid mutations in the vicinity of the active site are capable of shifting the product distribution towards the preference of every possible product. The underlying mechanism of this product formation control is still not conclusively resolved. Several computational approaches are applied to analyze possible steering factors for the selectivity in CspLOX2. Investigation of small radical substrate models, calculation of full active site structures as well as the simulation of the entire enzyme are carried out with varying levels of theory. The aim is to explore all factors which are potentially responsible for the enzyme selectivity and thus yield conclusive insights to the controlling mechanism of product formation in CspLOX2 and its mutants.

The structure of this thesis is organized as follows. Chapter 2 provides an overview on the methods and approximations applied in the context of this work. The theoretical background of quantum mechanical self-consistent field Hartree-Fock and Kohn-Sham density functional theory approaches is supplemented by the discussion of local orbital spaces and density fit-

ting approximations. Subsequently the classical counterpart to quantum mechanical treatments in form of molecular mechanics and molecular dynamics simulations are introduced. The chapter closes with insights to the typical tasks and issues regarding molecular fragmentation techniques on the example of two established procedures.

Chapter 3 deals with the use of Fock potentials in molecular embedding applications. In the first part the employment of such potentials in a cap-free fragment approach is presented and analyzed in terms of the possibility to recover the full molecular Hartree-Fock description from embedded monomer calculations. The scaling of the approach compared to the conventional Hartree-Fock procedure is evaluated and further effort is put into the continuous reduction of the involved computational cost. In a second part, application of the approach to the analysis of QM/MM embedding techniques is presented. This allows for the investigation of the importance to employ accurate potentials for the representation of a molecular environment.

In chapter 4 local density fitting approximations for the accelerated computation of Hartree-Fock exchange contributions are introduced and applied. Investigation of open-shell radical structures is carried out, providing insight into the limitations of the approximation and the actual speed up in self-consistent field approaches. Different formulations of Hartree-Fock and hybrid density functional theory are applied to analyze the performance of the local density fitting approximations in numerous procedures.

Chapter 5 contains the broad selectivity analysis of the CspLOX2 enzyme. The system is investigated by simulations of various size and accuracy. Starting from small model systems, the substrate structure is investigated quantum mechanically. For a large model of the complete active site the local density fitting approximations shown in chapter 4 are employed to investigate the substrate radical within the enzyme pocket. Molecular dynamics simulations are performed for the entire enzyme in solution to sample the conformational space of the bound substrate. In the end, a completely new procedure was developed to analyze the factor of steric shielding within the lipoxygenase system.

2 Theoretical Background

2.1 Quantum Mechanics

In quantum mechanics (QM) the state of a system is completely described by its wave function Ψ . This work will only consider time-independent problems when Ψ is a function of spatial and spin coordinates. Solving the corresponding Schrödinger equation

$$\hat{H}\Psi = E\Psi \quad (2.1)$$

yields the allowed states in a system defined by the Hamiltonian operator \hat{H} . The wave function Ψ is an eigenfunction of this operator with the corresponding eigenvalue of the energy E . By using atomic units, the Hamiltonian operator of a system consisting of n electrons and M nuclei can be written as

$$\begin{aligned} \hat{H} = & \underbrace{-\frac{1}{2} \sum_{i=1}^n \nabla_i^2}_{\hat{T}_e} - \underbrace{\frac{1}{2} \sum_{k=1}^M \frac{1}{m_k} \nabla_k^2}_{\hat{T}_n} - \underbrace{\sum_{k=1}^M \sum_{i=1}^n \frac{Z_k}{r_{ik}}}_{V_{en}} \\ & + \underbrace{\sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{r_{ij}}}_{V_{ee}} + \underbrace{\sum_{k=1}^M \sum_{l=k+1}^M \frac{Z_k Z_l}{r_{kl}}}_{V_{nn}}, \end{aligned} \quad (2.2)$$

where i and j are the indices of the electrons and k and l the ones of the nuclei. The mass and charge of a nuclei k is represented by m_k and Z_k , respectively. r_{xy} corresponds to the distance between two particles. The terms of equation (2.2) are divided into the kinetic energy of the electrons (\hat{T}_e) and the nuclei (\hat{T}_n), the potential energy of the attractive interaction between electrons and nuclei (V_{en}) and the potential energy of the repulsion of electrons (V_{ee}) or nuclei (V_{nn}).

However, this equation can be solved analytically only for very simple cases. In systems of chemical application one is forced to approximate the wave function Ψ and/or the Hamiltonian operator \hat{H} , in order to reduce the high-dimensionality of the problem. One of the most fruitful approaches in quantum chemistry, which is used in all of the methods to be discussed later on, is the Born-Oppenheimer approximation.^[12;13] In a system coupled by Coulomb interactions consisting of nuclei and electrons, both types of particles experience comparable forces leading to similar translational momenta. Due to the substantial mass difference the resulting velocity of the lighter electrons is significantly higher than the one of the nuclei, allowing for a separation of both movements. The problem can be decoupled

and split into two parts, namely the solution of an electronic Hamiltonian for fixed nuclear positions and the solution of a nuclear Hamiltonian within an effective electronic potential.

The energy of an electronic state for given nuclear positions is obtained by solving an electronic Schrödinger equation of the form

$$\hat{H}_{\text{el}}\Psi = E_{\text{el}}\Psi, \quad (2.3)$$

with the electronic Hamiltonian

$$\begin{aligned} \hat{H}_{\text{el}} &= -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{k=1}^M \sum_{i=1}^n \frac{Z_k}{r_{ik}} + \sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{r_{ij}} \\ &= \sum_{i=1}^n \hat{h}(i) + \sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{r_{ij}}. \end{aligned} \quad (2.4)$$

The missing nuclear repulsion potential is added to the electronic energy E_{el} *a posteriori*. Therefore, the further used terms "Schrödinger equation" and "Hamiltonian operator" implicitly reference to equation (2.3) and (2.4), respectively.

2.1.1 Hartree-Fock Theory

In the Hartree-Fock (HF) method, the wave function is represented by a single Slater determinant as

$$\Psi^{\text{HF}} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_n(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \psi_1(\mathbf{x}_n) & \cdots & \cdots & \psi_n(\mathbf{x}_n) \end{vmatrix}. \quad (2.5)$$

Here \mathbf{x}_i denotes a vectors containing the spatial and spin coordinates of an electron i as $\mathbf{x}_i = \{\mathbf{r}_i, \mathbf{s}_i\}$. The wave function Ψ^{HF} is an antisymmetric product of spin orbitals $\{\psi_i\}$, which in turn represent a product of a spatial orbital $\{\phi_i\}$ and a spin function α or β . The use of a Slater determinant ensures the Pauli principle for fermions to be fulfilled, which states that the exchange of two electrons only leads to a change of the sign of the wave function. In cases with an equal number of α and β spin electrons, the configuration can be restricted to occupation of the same spatial orbitals $\{\phi_i\}$ for one electron of each spin set. This is referred to as a *closed-shell restricted Hartree-Fock* (RHF) representation in which only the first $n/2$ orbitals are occupied. According to the Slater Condon

rules^[14], the expectation value for the closed-shell energy writes as

$$E_{\text{RHF}} = \langle \Psi^{\text{HF}} | \hat{H} | \Psi^{\text{HF}} \rangle = 2 \sum_{i=1}^{n/2} \langle i | \hat{h} | i \rangle + \sum_{i=1}^{n/2} \sum_{j=1}^{n/2} 2 \langle ii | jj \rangle - \langle ij | ji \rangle . \quad (2.6)$$

Here, the one-electron integrals take the form

$$\langle i | \hat{h} | j \rangle = \int \int \phi_i^*(\mathbf{r}_1) \hat{h}(\mathbf{r}_1) \phi_j(\mathbf{r}_1) \, d\mathbf{r}_1 , \quad (2.7)$$

while use has been made of a Mulliken-like notation for the two-electron integrals as

$$\langle ij | kl \rangle = \int \int \frac{\phi_i^*(\mathbf{r}_1) \phi_j(\mathbf{r}_1) \phi_k^*(\mathbf{r}_2) \phi_l(\mathbf{r}_2)}{r_{12}} \, d\mathbf{r}_1 d\mathbf{r}_2 . \quad (2.8)$$

For the energy expression (2.6), the integration over the spin coordinates has already been carried out, so that the following derivations are spin-free.

Hartree-Fock theory is by construction following the variational principle, according to which the expectation value of the Hamiltonian operator of any test wave function Ψ is always an upper bound to the exact energy E_0 of the system^[15]

$$\frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = E \geq E_0 .$$

This states a useful criterion for the optimization of the HF wave function. The lowering of the resulting energy value corresponds to a better description of the system under consideration. Thus, the optimal HF wave functions exhibits a minimum point of its energy expectation value with respect to orbital changes. Due to the side condition of orthonormal spatial orbitals, the optimization of the Hartree-Fock energy is achieved by the method of Lagrange multipliers. The quantity to be optimized is then the *Lagrange functional*

$$\mathcal{L} = E_{\text{RHF}} - 2 \sum_{i=1}^{n/2} \sum_{j=1}^{n/2} \epsilon_{ji} [\langle i | j \rangle - \delta_{ij}] , \quad (2.9)$$

with the Lagrange multipliers ϵ_{ji} and the overlap integral $\langle i | j \rangle$. By setting the derivative of this functional equal to zero for each orbital, one can derive the *Hartree-Fock equations*

$$\hat{f}(i) | i \rangle = \sum_{j=1}^{n/2} \epsilon_{ji} | j \rangle . \quad (2.10)$$

Within this, the *Fock operator* for each spatial orbital $\hat{f}(i)$ is written as

$$\hat{f}(i) = \hat{h}(i) + \sum_{j=1}^{n/2} \left[2\hat{j}_j(i) - \hat{k}_j(i) \right] = \hat{h}(i) + \hat{g}(i) . \quad (2.11)$$

The two operators $\hat{j}_j(i)$ and $\hat{k}_j(i)$ are named the *Coulomb* and *exchange* operator, respectively. They are defined via their effect on a spatial orbital:

$$\hat{j}_j(1)\phi_i(\mathbf{r}_1) = \phi_i(\mathbf{r}_1) \int \frac{|\phi_j(\mathbf{r}_2)|^2}{r_{12}} d\mathbf{r}_2 , \quad (2.12)$$

$$\hat{k}_j(1)\phi_i(\mathbf{r}_1) = \phi_j(\mathbf{r}_1) \int \frac{\phi_j^*(\mathbf{r}_2)\phi_i(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 . \quad (2.13)$$

By comparison of the Hamiltonian and the Fock operator in equations (2.4) and (2.11), the approximation made in Hartree-Fock theory becomes evident. The exact electron-electron operator r_{12}^{-1} is replaced by a mean-field operator $\hat{g}(i)$. Thus, in HF approaches each electron moves in an average potential of all remaining electrons.

It can be easily shown that the HF energy is invariant with respect to unitary transformations among the occupied orbitals. This enables an easier form of the HF equations (2.10) by a transformation that diagonalizes the Fock operator as

$$\hat{f}(i) |i\rangle = \epsilon_i |i\rangle . \quad (2.14)$$

The resulting spatial orbitals are then referred to as *canonical orbitals* with the corresponding orbital energies ϵ_i .

A crucial point to the solution of the canonical HF equations (2.14) is the construction of the spatial orbitals $\{\phi_i\}$. Since their exact form is unknown, approximations are needed to find a suitable representation of such orbitals. A common approach is use real molecular orbitals built from atom-like gaussian functions, the atomic orbital (AO) set. The *linear combination of atomic orbitals* (LCAO) approach incorporates these functions in an expanding series as

$$\phi_i(\mathbf{r}) = \sum_{\mu=1}^{N_{\text{bas}}} C_{\mu i} \chi_{\mu}(\mathbf{r}) . \quad (2.15)$$

In the limit of an infinite number of basis functions the expansion becomes exact and the correct wave function under the HF approximation is obtained. This is of course not feasible and the expansion always has to be restricted to a limited number of functions N_{bas} , which is referred to as the basis set size.

Introduction of the LCAO approach to the canonical HF equations (2.14) provides a matrix representation over the atomic orbitals in form of the

Hartree-Fock-Roothaan equations

$$\mathbf{FC} = \mathbf{SCE} . \quad (2.16)$$

Here, \mathbf{F} is the AO matrix representation of the Fock operator \hat{f} , \mathbf{C} the matrix of orbital coefficients and \mathbf{S} the AO overlap matrix with elements as $S_{\mu\nu} = \langle \mu | \nu \rangle$. The diagonal matrix \mathbf{e} contains the orbital energies. Application of the LCAO procedure to the energy expression (2.6) yields

$$\begin{aligned} E_{\text{RHF}} &= \sum_{\mu=1}^{N_{\text{bas}}} \sum_{\nu=1}^{N_{\text{bas}}} D_{\mu\nu} \left\{ h_{\mu\nu} + \frac{1}{2} \sum_{\rho=1}^{N_{\text{bas}}} \sum_{\sigma=1}^{N_{\text{bas}}} D_{\rho\sigma} \left[(\mu\nu | \rho\sigma) - \frac{1}{2} (\mu\sigma | \rho\nu) \right] \right\} , \\ &= \frac{1}{2} \text{tr} [\mathbf{D}(\mathbf{h} + \mathbf{F})] . \end{aligned} \quad (2.17)$$

For this expression the Fock matrix \mathbf{F} can also be expressed in terms of the one-electron matrix \mathbf{h} , the Coulomb matrix \mathbf{J} and the exchange matrix \mathbf{K} according to

$$\mathbf{F} = \mathbf{h} + 2\mathbf{J} - \mathbf{K} . \quad (2.18)$$

The one-electron density matrix \mathbf{D} is built from atomic orbital coefficients $C_{\mu i}$ as

$$D_{\mu\nu} = \sum_{i=1}^{n/2} C_{\mu i} C_{\nu i} \quad (2.19)$$

Since this sum only runs over the indices of the first $n/2$ molecular orbitals, the HF density and energy are purely defined by the occupied space. The remaining $N_{\text{bas}} - n/2$ orbitals are called *virtual orbitals*. They have no impact on the HF solution, but become of particular interest in post-HF treatments.

The Fock matrix \mathbf{F} in the basis of the atomic orbital functions, as given in equation (2.18), can be transformed to the Fock matrix \mathbf{f} in the basis of the molecular orbitals by an AO→MO transformation with the orbital coefficients as

$$\mathbf{f} = \mathbf{C}^\dagger \mathbf{F} \mathbf{C} . \quad (2.20)$$

This Fock matrix \mathbf{f} will be diagonal, if it is calculated with the canonical orbitals. Since the canonical orbital solution is not known from the start, the Fock matrix is at the beginning calculated with non-optimized orbitals. The first choice of coefficients to represent the orbital space is completely arbitrary and is often referred to as the *starting guess*. Via diagonalization of the Fock matrix \mathbf{f} access to a new set of molecular orbital coefficients is provided, following the extended eigenvalue problem of equation (2.16). These orbitals can be used to recompute the electronic density \mathbf{D} and from this once again another AO Fock matrix \mathbf{F} . AO→MO transformation

according to equation (2.20) leads to a new Fock matrix in the MO basis and the whole cycle can be started again. Thus, an iterative procedure is obtained which can be performed until self-consistency is reached and the involved molecular orbitals do not change anymore. Such an approach is referred to as the *self-consistent field* (SCF) method.

However, the SCF procedure in the MO basis applies an AO→MO transformation with the calculated coefficients \mathbf{C} in every iterative step. This may lead to an accumulation of numerical errors. A more convenient way to perform the optimization of the coefficients is provided in a symmetrically orthogonalized basis. The AO Fock Matrix and the coefficients are transformed with the AO overlap matrix \mathbf{S} according to

$$\tilde{\mathbf{F}} = \mathbf{S}^{-1/2} \mathbf{F} \mathbf{S}^{-1/2} , \quad (2.21)$$

$$\mathbf{C} = \mathbf{S}^{-1/2} \tilde{\mathbf{C}} . \quad (2.22)$$

This allows for the reduction of the generalized eigenvalue problem in equation (2.16) to a normal eigenvalue expression of the form

$$\tilde{\mathbf{F}} \tilde{\mathbf{C}} = \tilde{\mathbf{C}} \mathbf{e} . \quad (2.23)$$

Equation (2.23) provides the starting point for a numerically stable SCF procedure, in which the matrix $\tilde{\mathbf{F}}$ is diagonalized to yield the eigenvector matrix of the coefficients $\tilde{\mathbf{C}}$. Back transformation to the AO basis according to equation (2.22) yields access to a new Fock matrix \mathbf{F} and the next SCF iteration step can be carried out.

Up to this point, the case of an equivalent number of α and β electrons was assumed. The situation of a system with $n = n_\alpha + n_\beta$ electrons and $n_\alpha > n_\beta$ is referred to as an *open-shell* configuration. In this case the application of two different procedures is possible, which are described in the following paragraphs.

Restricted open-shell Hartree-Fock

In the formulation of *restricted open-shell Hartree-Fock* (ROHF) all n_β electrons of the system are assumed to be paired with an α electron in the same spatial orbital. This results in n_β doubly-occupied closed-shell and $(n_\alpha - n_\beta)$ open-shell orbitals, of which the latter are occupied by single α electrons. The split of the problem in open and closed parts affects the density matrix as

$$D_{\mu\nu} = 2 \sum_{i=1}^{n_\beta} C_{\mu i} C_{\nu i} + \underbrace{\sum_{t=1}^{n_\alpha - n_\beta} C_{\mu t} C_{\nu t}}_{D_{\mu\nu}^\alpha} . \quad (2.24)$$

Thus, the total density \mathbf{D} is divided into separate contributions. While the closed-shell part does not differ from the original density matrix, an additional term is introduced by the open-shell density matrix \mathbf{D}^o . This split of the density results in two different Fock matrices for both components as

$$F_{\mu\nu}^c = h_{\mu\nu} + \sum_{\rho=1}^{N_{\text{bas}}} \sum_{\sigma=1}^{N_{\text{bas}}} D_{\rho\sigma} \left[(\mu\nu|\rho\sigma) - \frac{1}{2}(\mu\sigma|\rho\nu) \right], \quad (2.25)$$

$$F_{\mu\nu}^o = -\frac{1}{2} \sum_{\rho=1}^{N_{\text{bas}}} \sum_{\sigma=1}^{N_{\text{bas}}} D_{\rho\sigma}^o (\mu\sigma|\rho\nu). \quad (2.26)$$

With these matrices the energy expression for the ROHF case is provided by distinct closed and open contributions:

$$E_{\text{ROHF}} = \frac{1}{2} \text{tr} [\mathbf{D}(\mathbf{h} + \mathbf{F}^c)] + \frac{1}{2} \text{tr} [\mathbf{D}^o \mathbf{F}^o]. \quad (2.27)$$

An AO \rightarrow MO transformation by analogy with equation (2.20) yields the closed- and open-shell Fock matrices in the MO basis, \mathbf{f}^c and \mathbf{f}^o . These matrices as well as their sum and difference can be incorporated in a block matrix $\bar{\mathbf{f}}$ according to^[16]

$$\bar{\mathbf{f}} = \begin{pmatrix} \mathbf{f}^c & \mathbf{f}^c - \mathbf{f}^o & \mathbf{f}^c \\ \mathbf{f}^c - \mathbf{f}^o & \mathbf{f}^c + \mathbf{f}^o & \mathbf{f}^c + \mathbf{f}^o \\ \mathbf{f}^c & \mathbf{f}^c + \mathbf{f}^o & \mathbf{f}^c + \mathbf{f}^o \end{pmatrix} = \begin{pmatrix} \mathbf{e}^c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{e}^o & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{e}^v \end{pmatrix}. \quad (2.28)$$

This block matrix can be used as an effective Fock matrix in the conventional SCF approach. It covers the different parts of the restricted open-shell formulation in form of the closed, open and virtual orbital spaces in its rows and columns. The matrices \mathbf{e} are diagonal and contain the energies of the closed (\mathbf{e}^c), open (\mathbf{e}^o) and virtual (\mathbf{e}^v) orbitals. Thus, diagonalization of $\bar{\mathbf{f}}$ provides access to the canonical orbitals of the ROHF formulation.

Unrestricted Hartree-Fock

Release of all restrictions on the spatial orbitals allows for the occupation of different orbital sets by α and β spin electrons. Within these sets all molecular orbitals are singly occupied and in general no closed-shell part is formed as seen in the previous paragraph. The procedure is then referred to as *unrestricted Hartree-Fock* (UHF) falling into the class of *different orbitals for different spins* (DODS) methods. This approach splits the corresponding problem into two subspaces, in which the SCF procedure is carried out for α and β spin electrons separately.

Two distinct density matrices result as

$$D_{\mu\nu}^{\alpha} = \sum_{i=1}^{n_{\alpha}} C_{\mu i}^{\alpha} C_{\nu i}^{\alpha} , \quad (2.29)$$

$$D_{\mu\nu}^{\beta} = \sum_{k=1}^{n_{\beta}} C_{\mu k}^{\beta} C_{\nu k}^{\beta} . \quad (2.30)$$

Their sum forms the total charge density matrix $\mathbf{D} = \mathbf{D}^{\alpha} + \mathbf{D}^{\beta}$. This allows for the formation of the α and β Fock matrices

$$F_{\mu\nu}^{\alpha} = h_{\mu\nu} + \sum_{\rho=1}^{N_{\text{bas}}} \sum_{\sigma=1}^{N_{\text{bas}}} [D_{\rho\sigma}(\mu\nu|\rho\sigma) - D_{\rho\sigma}^{\alpha}(\mu\sigma|\rho\nu)] , \quad (2.31)$$

$$F_{\mu\nu}^{\beta} = h_{\mu\nu} + \sum_{\rho=1}^{N_{\text{bas}}} \sum_{\sigma=1}^{N_{\text{bas}}} [D_{\rho\sigma}(\mu\nu|\rho\sigma) - D_{\rho\sigma}^{\beta}(\mu\sigma|\rho\nu)] . \quad (2.32)$$

Both Fock matrices are applied in HF-Roothaan eigenvalue expressions comparable to the one from closed-shell HF in equation (2.16):

$$\mathbf{F}^{\alpha} \mathbf{C}^{\alpha} = \mathbf{S} \mathbf{C}^{\alpha} \mathbf{e}^{\alpha} , \quad (2.33)$$

$$\mathbf{F}^{\beta} \mathbf{C}^{\beta} = \mathbf{S} \mathbf{C}^{\beta} \mathbf{e}^{\beta} . \quad (2.34)$$

Since the Fock matrices \mathbf{F}^{α} and \mathbf{F}^{β} both depend on the total charge density \mathbf{D} , the two equations are coupled and need to be solved simultaneously. The final UHF energy expression writes as

$$E_{\text{UHF}} = \frac{1}{2} \text{tr} [\mathbf{D}^{\alpha} (\mathbf{h} + \mathbf{F}^{\alpha})] + \frac{1}{2} \text{tr} [\mathbf{D}^{\beta} (\mathbf{h} + \mathbf{F}^{\beta})] . \quad (2.35)$$

In contrast to the restricted approaches, the UHF wave function is in general not an eigenfunction of the spin operator \hat{S}^2 . This arises from the artificial mixing of spins, which is referred to as *spin contamination* and can lead to serious problems with the description of the electronic state. Nevertheless, in some cases the unrestricted formalism is indispensable as the corresponding restricted formulation only converges to a saddle point instead of a true energy minimum. The removal of the applied restrictions and check for the corresponding UHF solution is a possibility to overcome such situations, which are often referred to as *triplet instabilities*.^[17]

The Hartree-Fock procedure provides an approximate solution to the Schrödinger equation with electrons moving in an averaged potential of their counterparts. This mean-field approach commonly covers 99% of the total electronic energy. The lacking 1% stands as the difference to the exact

energy and is defined as the *correlation energy* E_{corr} in

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} . \quad (2.36)$$

This contribution describes the instantaneous correlation of electrons during their movement, which can not be covered by an averaged electronic potential. However, the correlation effect is a key element of the electronic state and of crucial importance for the description of chemical transitions. The approaches covering this energy term are called *correlated methods*. Such methods often employ the Hartree-Fock solution as reference and evaluate the correlation contributions by a perturbative treatment or through the introduction of additional Slater determinants.

The canonical Hartree-Fock procedure formally scales as $\mathcal{O}(\mathcal{N}^4)$ with the molecular system size. This means doubling of the size of the investigated system will increase the computational time needed for the HF evaluation by a factor of 16. Possibilities for the reduction of this large computational effort are the central concern of this work and will be discussed in manifold approaches.

2.1.2 Orbital Localization

The canonical Hartree-Fock orbitals are in general widely spread over the entire calculated molecular system. This does not match the picture typically employed in chemistry, in which a small set of atomic orbitals are combined to form bonding, non-bonding and anti-bonding molecular orbitals which are very local in space. Since the orbitals resulting from Hartree-Fock theory are invariant to a unitary transformation, a representation close to the chemical picture can be found without changing the description of the system. By application of a unitary matrix \mathbf{U} to the canonical HF orbitals ϕ_i^{canon} , a set of localized molecular orbitals (LMOs) ϕ_i^{local} is obtained according to

$$|\phi_i^{\text{local}}\rangle = \sum_j^{N_{\text{occ}}} |\phi_i^{\text{canon}}\rangle U_{ji} = \sum_{\mu}^{N_{\text{bas}}} |\chi_{\mu}\rangle L_{\mu i} . \quad (2.37)$$

Here, \mathbf{L} corresponds to the localization matrix used for the application to the atomic basis functions χ_{μ} with $\mathbf{L} = \mathbf{C}\mathbf{U}$. The resulting orbitals ϕ_i^{local} are restricted to specific areas within the molecular system with only a limited number of AO functions significantly contributing to each of them.

A variety of procedures of orbital localization based on different localization matrices \mathbf{L} have been proposed. The choice of a specific operator to be optimized during the transformation distinguishes the approaches. While the popular localization procedure by Boys optimizes the spatial extent of the orbitals^[18], the Pipek-Mezey localization maximizes the sum

of orbital-dependent partial charges.^[19] Even though the latter approach ensures a clean separation of σ - and π -orbitals, it is commonly based on the use of Mulliken charges, which lack a mathematically sound definition.

Another fast and robust orbital localization technique has been introduced by Knizia in form of the intrinsic bond orbitals (IBOs).^[20] With a simple algebraic construction these localized orbitals provide insight into different aspects of chemical phenomena by separation into atomic core, valence and bond orbitals without touching the representation obtained by the self-consistent field wave function. Since the corresponding IBO partial atomic charges are nearly independent of the applied basis set size, the procedure stands as a stable orbital localization approach and will be applied within the context of this work.

2.1.3 Density Functional Theory

In Hartree-Fock theory the quantity of interest is the wave function. Under neglect of the electron spin, the evaluation of this quantity corresponds to a high-dimensional problem with $3n$ spacial coordinates. Density functional theory (DFT) provides a different approach for the description of the electronic state with a drastically reduced number of variables.

For the ground state of a molecular system, a one-to-one relation between the electron density $\rho(\mathbf{r})$ and the resulting energy E is pointed out by the Hohenberg-Kohn theorem.^[21] This enables a different approach to the problem stated by the Schrödinger equation. Instead of the search for an eigenfunction of the Hamiltonian operator, a functional connecting the electron density with the resulting energy is employed. This heavily reduces the dimensionality of the problem, since the electron density only depends on three spatial coordinates.

The general form of a spin independent density functional writes as

$$E[\rho(\mathbf{r})] = -\frac{1}{2} \sum_{i=1}^n \langle \phi_i | \nabla^2 | \phi_i \rangle + \int v(\mathbf{r}) \rho(\mathbf{r}) \, d\mathbf{r} + \frac{1}{2} \int \int \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d\mathbf{r} d\mathbf{r}' + E_{xc}[\rho(\mathbf{r})]. \quad (2.38)$$

In molecular systems, the external potential $v(\mathbf{r})$ is given by the potential of the nuclei, so that the second term respects electron-nuclei interactions. The third term takes into account electron-electron interactions via Coulombic repulsion of two electron clouds $\rho(\mathbf{r})$ and $\rho(\mathbf{r}')$. The exact form of the exchange-correlation functional $E_{xc}[\rho(\mathbf{r})]$ is unknown, which yields a variety of density functionals applying different approximations to the description of this term.

The first term of equation (2.38) respects the kinetic energy of the electrons. However, in contrast to the other contributions it is not constructed

as a functional of the density, but rather via the use of molecular orbitals $\{\phi_i\}$. A density-based description of the kinetic energy term leads to significant errors. Indeed, in an orbital-free formulation of this functional, no chemical system would be stable against its dissociation.

Introduction of molecular orbitals in DFT approaches is referred to as the *Kohn-Sham* (KS) procedure. Even though this procedure extends the dimension of the problem to $3n$ spatial coordinates, it results in solving equations close to the canonical HF equations (2.14) for each molecular orbital:

$$\underbrace{\left[-\frac{1}{2}\nabla^2 + v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{ex}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \right]}_{\hat{f}^{\text{KS}}} |\phi_i\rangle = \epsilon_i \phi_i . \quad (2.39)$$

Here, the electron density is expanded in terms of the molecular Kohn-Sham orbitals by

$$\rho(\mathbf{r}) = 2 \sum_{i=1}^n \phi_i^* \phi_i . \quad (2.40)$$

As already seen for Hartree-Fock theory, the molecular orbitals can be expressed by atomic basis functions through application of the LCAO approach of equation (2.15). Thus, a matrix representation of the Kohn-Sham operator \hat{f}^{KS} in equation (2.39) is possible and a generalized eigenvalue relation results, which is identically in form to the Hartree-Fock-Roothaan equations (2.16):

$$\mathbf{F}^{\text{KS}} \mathbf{C} = \mathbf{S} \mathbf{C} \epsilon . \quad (2.41)$$

Here, \mathbf{F}^{KS} is closely related to the conventional Fock matrix and is referred to as the Kohn-Sham matrix. This exhibits that the Kohn-Sham procedure can also be performed in an iterative way by application of the SCF approach. It employs the same treatment of atomic and molecular orbitals and therefore the KS approach can also be performed in restricted closed-shell (RKS), restricted open-shell (ROKS) and unrestricted (UKS) formulations. According to this, the expression of a self-consistent field procedure is from now on used as a generic term for Hartree-Fock and Kohn-Sham DFT methods.

As already indicated above, the crucial part of a DFT calculation is the definition of the exchange-correlation functional, since no analytic expression is available. The large variety of density functionals available in the literature employ approximations of different quality to the formation of $E_{xc}[\rho(\mathbf{r})]$. The ranking of the involved approximations according to the complexity of the description is referred to as the *Jacob's ladder* of density functionals.^[22] On the lowest rung of this ladder, the local density approximation (LDA) is found. Corresponding approaches only use local

information on the density, by employing the values of $\rho(\mathbf{r})$ at certain points in space. The generalized gradient approximation (GGA) additionally takes into account the first derivative of the electron density and is thus located higher in the density functional ranking. Further improvement can be achieved, if also the second derivative is included, whereas the corresponding type of density functionals is then referred to as meta-GGA.

The next step on the ladder is the inclusion of non-local effects from other approaches, which are not directly depending on the density and its derivatives. Hybrid density functionals form the exchange part of $E_{xc}[\rho(\mathbf{r})]$ by admixture of exact exchange contributions from Hartree-Fock theory. This helps with the reduction of problems commonly observed in standard DFT calculations like the self-interaction error.^[23] An even higher rung of the ladder is represented by doubly hybrid functionals, which additionally incorporate correlation contributions from correlated methods. However, this last step is comparably costly and does not ensure a further improvement of the DFT result.

Nowadays by far the most popular density functional is the B3LYP functional.^[24;25] It belongs to the class of hybrid functionals and forms the exchange-correlation contribution according to

$$E_{xc}^{\text{B3LYP}} = E_x^{\text{LDA}} + a_0 (E_x^{\text{HF}} - E_x^{\text{LDA}}) + a_x (E_x^{\text{GGA}} - E_x^{\text{LDA}}) + E_c^{\text{LDA}} + a_c (E_c^{\text{GGA}} - E_c^{\text{LDA}}) . \quad (2.42)$$

The starting point is the LDA approximation E_x^{LDA} from Slater-Dirac exchange.^[26] This term is corrected by the addition of Hartree-Fock exchange E_x^{HF} , controlled by the first parameter a_0 . Another correction to the exchange part is the admixture of exchange contributions resulting from the GGA approximation E_x^{GGA} taken from the Becke88 functional.^[27] This contribution is adjusted with the second parameter a_x . The initial contributions to the correlation part are extracted from LDA approximations E_c^{LDA} in the Lee-Yang-Parr functional.^[28] This is finally corrected in terms of the third parameter a_c , with the correlation contributions E_c^{GGA} from the VWN80^[29] functional under GGA approximation. The parameters in B3LYP are set to $a_0 = 0.2$, $a_x = 0.72$ and $a_c = 0.81$ resulting from fits to optimally reproduce chemical properties of the G1 molecule test set.

Construction of a universal density functional yielding accurate results for every system under consideration is desirable, but difficult due to the parametrization of the exchange-correlation functional. A large list of functionals have been proposed to accurately describe a broad variety of chemical phenomena. However, in most cases these functionals perform best for the type of systems they were parametrized to, while the resulting issues in applications to other problems can not be foreseen.

A common deficit of DFT approaches is the lacking description of dis-

persion interactions. These contributions arise from non-overlapping densities and thus can not be included by a simple density functional term. The approach of parametrized dispersion corrections is widely used to account for the missing interactions of the DFT functionals. In this context well-established is the D3 procedure introduced by Grimme^[30], which includes dispersion interactions by addition of simple pair-wise damped Lennard-Jones potentials to the resulting DFT energy.^[31]

2.1.4 Density Fitting

The evaluation of four-center electron-repulsion integrals (ERIs) is a computationally demanding task, but a fundamental operation in atomic orbital integral-based programs. In the case of Hartree-Fock theory this type of integrals arise in form of the Coulomb and exchange contributions to the Fock matrix. In density functional theory ERIs over the Kohn-Sham orbitals are found regarding the Coulomb electron-electron interactions. Since hybrid DFT functionals employ the exact exchange contributions from HF, the corresponding calculations lead to the computation of four-index exchange integrals as well. Several approaches were proposed to accelerate the evaluation of the Fock/Kohn-Sham matrix contributions by application of screening techniques to the demanding ERIs and the elements of the density matrix. Especially for the more rapidly decaying exchange terms these methods enabled a quadratic or even linear scaling of the corresponding evaluation with growing system size.^[32;33] However, the original scaling of the procedure with an increasing size of the applied basis set remained unchanged. A popular approach proposed by Neese and coworkers overcame this problem by the use of a chain of spheres exchange (COSX)^[34–36], which finds an approximate exchange potential by performing a seminumerical integration of the four-center ERIs.

In density fitting (DF)^[37] approaches a different idea is followed to find approximate descriptions of the Coulomb and exchange integrals. An auxiliary basis set is introduced to represent the four-center ERIs as combinations of two- and three-index integrals, similar to methods based on the resolution of the identity (RI)^[38–41] and Cholesky decomposition.^[42] In the following a notation is applied to discriminate the involved basis functions and orbitals. Atomic basis functions are labeled by Greek letters $\{\mu, \nu, \dots\}$, molecular orbitals as seen before by small Latin letters $\{i, j, \dots\}$ and the auxiliary fitting functions by capital letters $\{A, B, \dots\}$.

All two-electron integrals over the atomic basis functions can be expressed as electrostatic interactions between two orbital-product densities $\rho_{\mu\nu}(\mathbf{r}) = \chi_{\mu}(\mathbf{r})\chi_{\nu}(\mathbf{r})$ as

$$(\mu\nu|\rho\sigma) = \int \int \frac{\rho_{\mu\nu}(\mathbf{r}_1)\rho_{\rho\sigma}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 . \quad (2.43)$$

The one-electron densities are approximated by the linear expansion

$$\rho_{\mu\nu}(\mathbf{r}) \approx \tilde{\rho}_{\mu\nu}(\mathbf{r}) = \sum_{A=1}^{N_{\text{fit}}} d_A^{\mu\nu} \chi_A(\mathbf{r}), \quad (2.44)$$

in which $\chi_A(\mathbf{r})$ are the functions of the auxiliary density fitting basis set. The involved coefficients of the expansion, $d_A^{\mu\nu}$, are obtained by minimization of the difference to the exact one-electron densities in the quantity $\Delta_{\mu\nu}$ as

$$\Delta_{\mu\nu} = \int \int \frac{[\rho_{\mu\nu}(\mathbf{r}_1) - \tilde{\rho}_{\mu\nu}(\mathbf{r}_1)][\rho_{\mu\nu}(\mathbf{r}_2) - \tilde{\rho}_{\mu\nu}(\mathbf{r}_2)]}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2. \quad (2.45)$$

This type of fitting minimizes the least squares error of the electric field^[43] and yields the expansion coefficients

$$d_B^{\mu\nu} = \sum_{A=1}^{N_{\text{fit}}} (\mu\nu|A) [\mathbf{Q}^{-1}]_{AB}, \quad (2.46)$$

where the elements of the fitting matrix Q_{AB} and the three-index integral $(\mu\nu|A)$ are defined as

$$Q_{AB} \equiv (A|B) = \int \int \frac{\chi_A(\mathbf{r}_1)\chi_B(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2, \quad (2.47)$$

$$(\mu\nu|A) = \int \int \frac{\chi_\mu(\mathbf{r}_1)\chi_\nu(\mathbf{r}_1)\chi_A(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2. \quad (2.48)$$

Subsequent substitution of the fitting relations (2.44), (2.48) and (2.46) in equation (2.43) enables an approximate description of the initial four-index two-electron integrals by

$$\begin{aligned} (\mu\nu|\rho\sigma) &= \int \int \frac{\rho_{\mu\nu}(\mathbf{r}_1)\rho_{\rho\sigma}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \\ &\approx \int \int \sum_{B=1}^{N_{\text{fit}}} \frac{d_B^{\mu\nu} \chi_B(\mathbf{r}_1)\rho_{\rho\sigma}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \sum_{B=1}^{N_{\text{fit}}} d_B^{\mu\nu} (B|\rho\sigma) = \sum_{A=1}^{N_{\text{fit}}} \sum_{B=1}^{N_{\text{fit}}} (\mu\nu|A) [\mathbf{Q}^{-1}]_{AB} (B|\rho\sigma). \end{aligned} \quad (2.49)$$

Equation (2.49) contains the core idea of the density fitting approximation. The four-center ERIs $(\mu\nu|\rho\sigma)$ are approximated by combinations of the two-index integrals of the fitting matrix entries $Q_{AB} = (A|B)$ and three-index integrals of the $(\mu\nu|A)$ type.

Application of the density fitting approach to the Coulomb matrix in a

closed-shell HF calculation yields

$$J_{\mu\nu} \approx \tilde{J}_{\mu\nu} = \sum_{A=1}^{N_{\text{fit}}} e_A (A|\mu\nu) , \quad (2.50)$$

where the coefficients e_A are determined by

$$e_A = \sum_{B=1}^{N_{\text{fit}}} \left[[\mathbf{Q}^{-1}]_{AB} \sum_{\mu=1}^{N_{\text{bas}}} \sum_{\nu=1}^{N_{\text{bas}}} D_{\mu\nu}(\mu\nu|B) \right] . \quad (2.51)$$

Here, $D_{\mu\nu}$ is the closed-shell electron density matrix as given in equation (2.19). Computation of the contributions to the exchange matrix is the more demanding part of the density fitting approach. The exchange matrix can be approximated as

$$K_{\mu\nu} \approx \tilde{K}_{\mu\nu} = \sum_{A=1}^{N_{\text{fit}}} \sum_{i=1}^{N_{\text{occ}}} (\mu i|A) d_A^{\nu i} , \quad (2.52)$$

where the original three-index integrals and expansion coefficients have been half-transformed with the orbital coefficients according to

$$(\mu i|A) = \sum_{\nu=1}^{N_{\text{bas}}} C_{\nu i}(\mu\nu|A) , \quad (2.53)$$

$$d_A^{\nu i} = \sum_{\mu=1}^{N_{\text{bas}}} C_{\mu i} d_A^{\mu\nu} = \sum_{B=1}^{N_{\text{fit}}} (\nu i|B) [\mathbf{Q}^{-1}]_{AB} \quad (2.54)$$

The one-electron matrix \mathbf{h} , the density fitted Coulomb matrix $\tilde{\mathbf{J}}$ and density fitted exchange matrices $\tilde{\mathbf{K}}$ are finally used to construct the the approximate closed-shell Fock matrix in the AO basis according to equation (2.18):

$$\tilde{F}_{\mu\nu} = h_{\mu\nu} + 2\tilde{J}_{\mu\nu} - \tilde{K}_{\mu\nu} . \quad (2.55)$$

In the case of a restricted open-shell HF (ROHF), the fitting of Coulomb and exchange contributions is performed separately for the closed and open Fock matrices \mathbf{F}^c and \mathbf{F}^o as given in equations (2.25) and (2.26), respectively. This allows for their final combination in the effective Fock matrix \mathbf{f} in equation (2.28). For unrestricted HF (UHF) the procedure is applied independently to the \mathbf{F}^α and \mathbf{F}^β Fock matrices given in equations (2.31) and (2.32) arising from the distinct treatment of α and β spin electrons.

The density fitting approximation is not restricted to SCF methods like HF and DFT. Any quantum chemical calculation that involves the

computation of four-index-integrals can benefit from the above-described procedure. It is commonly used in post-HF correlation methods like Møller-Plesset perturbation^[44] and coupled cluster theory^[45], in which the electron correlation is estimated by sets of two-electron four-index integrals.

Despite the introduced approximations the DF-HF method formally still scales as $\mathcal{O}(\mathcal{N}^4)$ with the molecular system size. However, the prefactor is drastically reduced and the rise of the computational cost with the size of the applied basis set becomes less demanding. This enables the calculation of medium-sized systems in reasonable computational times.

2.2 Fragmentation Techniques

Fragmentation approaches provide a straightforward way to reduce the computational effort of any type of molecular calculation by splitting up the system under consideration into smaller entities. The energy and other properties of the original structure are subsequently restored from the individual fragment calculations.^[7;8] This raises a set of issues which are of critical importance to the success of any fragmentation technique.

At first, the choice of the generated fragments may result in radical or even non-chemical molecules. Since these structures significantly differ from their state in the original environment, special treatment of the fragments is necessary allowing for the representation as a piece of the total molecular system. An approach often performed in this context is the introduction of molecular caps or link-atoms to saturate dangling bonds and form closed-shell fragments.

Another problem that arises from the separate treatment of molecular entities is the inclusion of their inter-fragment interactions. In an ideal case one would like to describe each fragment individually but under the interaction effect of the entire remaining system. This is the starting point for embedding techniques, in which the impact of the environment on a molecular calculation is included via an external potential. The most prominent potential of this type is the application of simple atomic point charges to account for the involved Coulomb interactions. A different way to address the issue of lacking inter-fragment interactions is their pair-wise incorporation from dimer calculations in two-body approaches. Through the combined treatment of all pairs of fragments and subsequent subtraction of the involved monomer energies, the interaction of the entities can be calculated in a straightforward manner. This idea can be developed in a series of higher-order contributions including many-body effects by subsequent calculations of molecular supersets like dimers, trimers, tetramers and so forth.

If quantum mechanical calculations are performed for the fragments, an ubiquitous complication arises from the treatment with a molecular basis set of limited size. The basis set superposition error (BSSE) describes the

fact that in the calculation of the total system each entity is artificially stabilized compared to the single fragment treatment due to the additional basis functions provided in the environment.^[46;47] This error vanishes in the limits of a complete basis set treatment. Several approaches to cancel BSSE effects have been discussed over the years, of which the counterpoise correction is by far most popular one.^[48;49] In this correction the artificial stabilization is completely removed by employing the full molecular basis set of the entire system in each fragment calculation. It is also often applied to the calculation of molecular dimer properties, in which the involved monomer calculations suffer from the same lack of basis set description compared to their combined treatment.

The issues regarding molecular fragmentation techniques are discussed in further detail on the example of two different established fragmentation methods in the course of this section.

2.2.1 Molecular Fractionation with Conjugate Caps

The method of molecular fractionation with conjugate caps (MFCC) introduced by Zhang and Zhang^[50] provides a fragmentation technique for the calculation of peptide and protein structures. It was originally proposed for the computation of protein-ligand interaction energies.^[51;52]

In the MFCC approach a protein or peptide is split into single amino acid fragments by cutting all peptide bonds. Both ends of each fragment are saturated by introduction of complete atom groups, the molecular caps. An acetyl group is added to the amino end of the fragment, while the carboxyl term is saturated with an *N*-methyl group. The introduced atomic groups ensure a representation of the natural environment of the amino acid within the protein and form a pair of conjugate caps (concaps, cc). This property can be exploited by combination of both caps in one molecule of *N*-methylacetamide. Subsequent subtraction of the calculation of this molecule allows for a clean cancellation of the cap contributions. Formation of the fragments and the resulting conjugate caps is shown on the example of a tetrapeptide in Figure 2.1.

The interaction energy between a ligand molecule *L* and a protein *P* which is decomposed into *N* fragments writes as

$$E_{P-L} = \sum_{i=1}^N E(F_i + L) - \sum_{i=1}^N E(F_i) - \sum_{i=1}^{N-1} E(cc_i + L) + \sum_{i=1}^{N-1} E(cc_i) - E(L). \quad (2.56)$$

Thus, the approach sums up all individual interactions between the fragments and the ligand molecule and subsequently cancels the contributions of the conjugate caps. Subtraction of the self-energies allows for the calculation of the interaction energy of the entire protein with the ligand and for the identification of amino acids with significant contribution to this

energy term. However, this approach does neither include the interactions of the fragments among each other, nor provide a way to calculate the total energy of the protein.

The problem of the lacking total energy description of the protein was overcome by introduction of the generalized MFCC approach (GMFCC).^[53] This expansion of the original method includes inter-fragment interactions at the level of an MM force field. Since this is a very rough description for

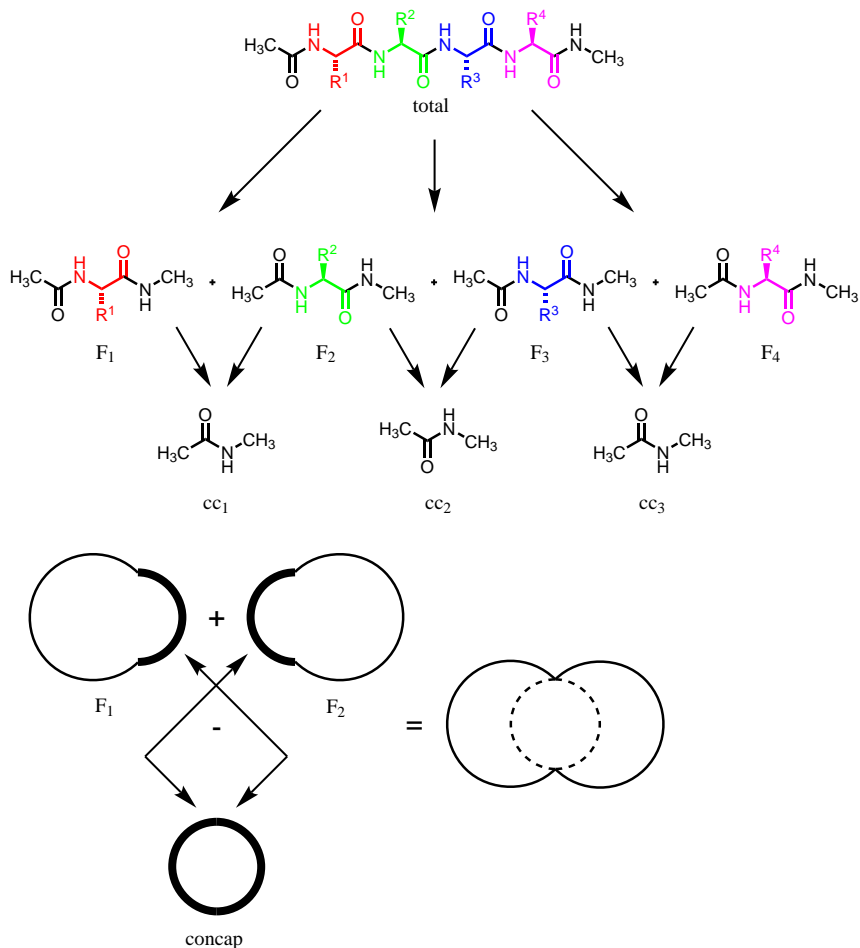


Figure 2.1: Fragmentation of the MFCC approach on the example of a tetrapeptide. Four fragments containing one amino acid each are constructed together with three conjugate caps. The bottom part illustrates the application of conjugate caps by combination of the introduced molecular groups to cancel their effect.

directed interactions, present in protein structures in form of salt bridges and hydrogen bonds, the approach also includes specific dimer contributions. Fragments that are non-neighboring within the peptide chain, but found to be close in space, are treated in a combined quantum mechanical calculation. The approach was even extended further by inclusion of point charges for the embedding of each fragment. This electrostatically embedded GMFCC (EE-GMFCC) was used to accurately reproduce protein energies with quantum mechanical fragment treatment.^[54]

Overall, the MFCC class of approaches ensures a strict linear scaling with growing size of the investigated protein, since the number of generated fragments increases constantly. Saturation of all fragments with molecular caps introduces no constraints to the level of theory applied. Use has been made of HF, DFT and post-HF correlation methods in MFCC approaches, providing linear-scaling procedures for the calculation of protein properties.^[55]

However, the proposed conjugate caps are limited to protein and peptide structures. A generalization of the idea of conjugate caps can be seen in the extended ONIOM scheme.^[56] Even though this approach does not employ molecular fragmentation, it follows the basic concept of conjugate caps by calculation of overlapping subsections of a system. Very much in the spirit of the MFCC method, the undesired contributions are subsequently canceled by subtraction of the intersection.

2.2.2 Fragment Molecular Orbital Method

The inclusion of intramolecular interactions in fragment-based approaches is a key element to the recovery of the original molecular state. A common attempt is the calculation of two-body terms in the electrostatic potential of the remaining system to account for higher-order interaction effects. Perhaps the most successful approach of this type is the fragment molecular orbital method (FMO), originally proposed by Kitaura and coworkers.^[57] Within this method the fragment calculations are converged in monomeric Coulomb potentials and many-body calculations are performed in the resulting field.

Every FMO procedure starts with free monomer calculations to obtain the initial electronic densities. By application of Coulomb operators of the remaining fragments the monomers calculations are subsequently iterated until self-consistency is reached. Non-iterative dimer calculations are performed within the obtained Coulomb bath, referred to as FMO2. The corresponding energy expression is given by the classical two-body expansion as

$$E_{\text{FMO2}} = \sum_{I=1}^N E_I + \sum_{I=1}^N \sum_{J=I+1}^N (E_{IJ} - E_I - E_J) . \quad (2.57)$$

Here, E_I and $E_{I,J}$ are the energies of the embedded monomers and dimers, respectively. Additional three- and four-body terms can be optionally carried out in higher-order FMO procedures referred to as FMO3 and FMO4.

In contrast to other fragmentation techniques, the FMO approach breaks covalent bonds by heterolytic fission. Thus, both bonding electrons are transferred to the same fragment. A resulting charge is avoided by a special treatment with an additional proton shift from the electron-lacking to the electron-accumulating fragment yielding two uncharged monomers. The frontier orbitals previously contributing to the covalent bond are frozen in the course of the fragment calculations to maintain the original character within the molecular system.^[58]

The choice of molecular fragmentation in FMO avoids the appearance of non-physical contributions from link-atoms or molecular caps. Due to the pair-wise treatment of all fragments within the dimer calculations, significant contributions to the intramolecular interaction can be easily detected^[59], which commonly employed for graphical representation of specific interactions on an electrostatic potential surface.^[60] However, the description is lacking contributions beside the embedding Coulomb potential of the fragments. The inclusion of the additional effects of exchange interactions to FMO has been discussed, but lead to unsatisfying results due to artifacts in the description of the dimers in the obtained monomer field.^[61]

The overall computational scaling of the FMO method is dominated by the quadratic growth of dimer calculations with the number of generated monomers. In addition, to incorporate the embedding Coulomb potentials and remove BSSE effects, all calculations are carried out with the full molecular basis set leading to a drastic increase of the computational effort. This issue of steep computational scaling can be overcome by exploiting the independence of the fragments calculations within the monomeric iteration cycles. Since no further information is needed for a specific fragment until the running monomer cycle is completed, the procedure is perfectly suited to be performed in parallel on several CPUs. The application of the FMO approach on supercomputer clusters allows for the routine calculation of large molecular structures with an accurate quantum mechanical description.

2.3 Molecular Mechanics

The aforementioned methods and approaches allow for an approximate solution of the Schrödinger equation. But as already stated, the computational effort of these descriptions scales exponentially with an increasing system size making them inapplicable for very large systems. Especially in the field of biomolecular computations one is quickly dealing with systems of

thousands of atoms ruling out any kind of quantum mechanical treatment.

An alternative is provided by molecular mechanics (MM) approaches, which describe the system in a classical fashion comparable to the principle of a molecular building kit. Within this simplified treatment atoms are considered as hard spheres and bonds between them are represented by springs. With the application of fitted parameters all interactions between the atoms are described by analytic functions of simple form. The evaluation of such functions is of extremely low computational cost, making the application to systems of several thousands of atoms amenable.

Incorporation of the classical motion of the nuclei in molecular dynamics (MD) simulations provides insights into the dynamical properties of the system under consideration. Due to the simplicity of the underlying equations external influences like solvation, temperature and pressure can be modeled without particular difficulties. This allows for manifold applications like the investigation of periodic solid state structures or the conformational sampling of biomolecules in their natural environment. The underlying potentials in form of force fields and the characteristics of molecular dynamics simulations are discussed in detail within this section.

2.3.1 Force Fields

The energy of the system in a molecular mechanical approach is calculated by parametrized classical potentials combined in a force field. Since electrons are not explicitly included, the total energy is given as a pure function of the nuclei coordinates. The parameters involved can be either fitted to experimental data or extracted from quantum mechanical model calculations. Every atom is assigned a specific atom type, defining the number of binding partners and the parameters used in the corresponding force field terms. A set of analytic functions is used to incorporate specific molecular contributions. The AMBER force field^[62], for example, uses the following energy terms:

$$\begin{aligned}
 V_{\text{AMBER}} = & \sum_{\text{bonds}} \frac{1}{2} k_b (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{torsions}} \frac{1}{2} k_\varphi [1 + \cos(n\varphi - \gamma)] + \sum_{\text{improper}} \frac{1}{2} k_\chi (\chi - \chi_0)^2 \\
 & + \sum_{i=1}^M \sum_{j=i+1}^M \frac{q_i q_j}{4\pi\epsilon_o\epsilon_r r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]. \quad (2.58)
 \end{aligned}$$

The first two terms represent the contributions of bonds and angles by harmonic potentials. Involved parameters are the force constants k_b and k_θ , as well as the equilibrium distance values of the corresponding internal

coordinate b_0 and θ_0 . Such functions provide reasonable approximations to the actual covalent contributions as long as the nuclei configuration does not deviate strongly from the equilibrium geometry. The introduction of harmonic potentials goes along with the limitation that covalent bonds can not be broken within this description.

In the third sum the contributions of dihedral or torsional angles are respected by Fourier series. Additional to another force constant k_φ , the parameter for the number of maxima within a 360° twist of the dihedral in form of the periodicity n and the phase shift γ are needed. Improper torsions are included in the fourth term of the MM potential and ensure the maintenance of planarity in specific structural motifs like phenyl rings or peptide bonds.

Non-bonding interactions are treated by the two last terms. Contributions are added up by a double sum to respect the Coulomb and van der Waals interactions of all pairs of atoms with a distance r between them. While Coulomb's law is followed for two point charges q_1 and q_2 located at the atoms in the first term, their van der Waals interaction is represented by a Lennard-Jones potential in the second. The latter requires two further parameters for each atom pair in form of the depth of the potential well ϵ and the zero point of the potential σ .

The simple form of the contributing potentials allow for the routine application of MM approaches to several thousands of atoms. A price to pay for this favorable scaling is the incapability of MM simulations to break or form covalent bonds. Since the connectivity of each atom has to be explicitly specified and does not change during the simulation, no reactive chemical transformation of the investigated system can be observed.

2.3.2 Molecular Dynamics

Molecular Dynamics (MD) simulations apply the classical Newton's equations of motion to obtain atom movements from an arbitrary potential energy expression. For a particle i with the mass m_i moving along one coordinate \mathbf{r}_i , Newton's second law requires solving of the differential equation

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{F_{\mathbf{r}_i}}{m_i}. \quad (2.59)$$

Here, $F_{\mathbf{r}_i}$ is the force acting on the particle in the direction of the movement. Different algorithms have been established for the integration of equation (2.59), by making use of a Taylor expansion for the position at a time $t + \delta t$

and its derivatives like the velocities $\mathbf{v}(t + \delta t)$ and accelerations $\mathbf{a}(t + \delta t)$ as

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) + \frac{1}{6} \delta t^3 \mathbf{b}(t) + \dots \quad (2.60)$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \delta t \mathbf{a}(t) + \frac{1}{2} \delta t^2 \mathbf{b}(t) + \dots \quad (2.61)$$

$$\mathbf{a}(t + \delta t) = \mathbf{a}(t) + \delta t \mathbf{b}(t) + \dots \quad (2.62)$$

The solution provides insights into the trajectory of the particle following its movement after certain periods of time, whereas the chosen difference δt is referred to as the *time step* of the simulation. Generalization of the procedure can be easily performed to allow for the application to a large number of particles. The approach is thus perfectly suited for the investigation of the dynamical properties of a chemical system, with being atoms treated as individual particles moving in the potential of the applied force field. Also combined movements of atoms are taken into account, according to which parts of the system or the entire simulated structure is moving in molecular motions.

Several ensembles can be simulated employing MD approaches by restriction of specific system variables. In the *microcanonical ensemble*, the amount of substance N , the volume V and the energy E of the system are conserved. The corresponding simulation is therefore also referred to as *NVE* dynamics. It allows for the redistribution of kinetic and potential energy, while the total energy of the system is kept constant.

The *canonical ensemble* also fixes the variables of N and V , but instead of the energy the temperature T is not allowed to vary anymore (*NVT*). This is achieved by thermostat, which commonly scales velocities of the moving atoms to obtain a specific temperature. Different approaches are present to achieve this constraint. The Berendsen thermostat for example couples the system to an external temperature bath.^[63] Introduction of friction effects to the applied Newton's equations of motion is the method of choice in case of the Langevin thermostat.^[64]

In the *isothermal-isobaric ensemble* the constrained quantities are the particles N , the temperature T and the pressure p (*NPT*). This comes closest to the experimentally accessible situation of ambient temperature and pressure. Constancy of the pressure is commonly achieved by scaling the volume of the system. The coupling to a pressure bath can be performed in analogy with the temperature fixation in the Berendsen thermostat.

In condensed phase simulations, one is usually interested in the simulation of a solute in a continuous solvent environment or the representation of an infinite crystal structure. Therefore, a special treatment is applied to overcome the limited size of the simulated system. By introduction of a periodic box, the system repeats itself after a certain distance in space

canceling the impact of surface effects. A cutoff is applied to restrict long-range interactions to a specific distance in order to decrease the computational effort. This type of simulations is referred to as being carried out under *periodic boundary conditions*.

3 Fock Potentials in Embedded Calculations

The Hartree-Fock method is at its core an embedding procedure. It provides a solution to the Schrödinger equation through a mean-field description of electrons moving in an effective electronic potential. Following this basic concept in systems of several monomers, it is possible to recover the full HF description from monomer calculations embedded in an effective mean potential of their counterparts. An earlier study by Mata and Stoll^[65] on non-covalently bonded systems showed that the use of full Fock potentials is suitable for an efficient monomer embedding. The Fock potential of a single monomer covers all of its information within the system at the HF level. Due to the mean-field character of the HF method itself, this type of embedding is by no means an approximation and the monomer calculations will reach the complete HF treatment in the limit of exact potentials used. Thus, the application of full Fock potentials provides a stable reference and can be used to critically analyze other established embedding techniques. The exact limit is reached by updates of the monomer Fock potentials in an iterative procedure, very much in the spirit of the updating SCF steps within the HF method. If such iterations are carried out to self-consistency, the resulting Fock potentials represent the converged description of the corresponding monomers within the full system.

This chapter gives an overview on the application of full Fock potentials in embedded calculations of molecular systems. A general introduction to the inclusion of Fock potentials via the use of effective one-electron operators is followed by the discussion of two different schemes. The first scheme introduces a cap-free fragment approach to HF calculations for covalent systems. After the detailed description of the method and procedure, the comparison of density and energy convergence to full HF results is presented. The analysis also considers the scaling of the fragment approach with the system size in descriptions with limited basis sets.

In the second scheme the application of QM/MM embedding techniques is analyzed by comparison to full Fock embedding. The investigation includes the effect on the monomer density of crystal structures and molecules in solution considering the type of embedding used for their first and second surrounding molecular shells. Analysis of the embedding effect on a higher-order molecular property is conducted on the example of the dispersion interaction energy of molecular dimers. A conclusion critically summarizing the benefits and limitations of embedding calculations with full Fock potentials closes the chapter.

3.1 One-electron Operators

In restricted closed-shell Hartree-Fock theory, including an effective potential of neighboring fragments to the calculation of a fragment x can be

expressed by changing the effective one-electron operator as

$$\mathbf{h}^x = \mathbf{h}_0^x + \sum_{y \neq x}^N [2\mathbf{J}^y - \mathbf{K}^y + \mathbf{V}_n^y + \mathbf{P}^y], \quad (3.1)$$

where \mathbf{J}^y and \mathbf{K}^y are the Coulomb and exchange matrices, respectively, of a surrounding monomer y and \mathbf{V}_n^y the corresponding nuclei potential. \mathbf{h}_0^x refers to the one-electron matrix of the free fragment, calculated without embedding. Thus, the embedding potential includes all contributions to the Fock matrix of the monomer y plus an additional term involved in form of the matrix \mathbf{P}^y . This term is the matrix representation of a shift operator \hat{p}^y defined as

$$\hat{p}^y = \lambda \sum_{i \in y}^{n_{\text{occ}}^y} |\phi_i\rangle \langle \phi_i|. \quad (3.2)$$

Here, ϕ_i represent the occupied molecular orbitals of the embedding fragment y , while λ is a constant with the dimension of an energy that should be set to high values. Since the energy is optimized during the SCF procedure, these energetically high terms are avoided by minimizing the overlap between the orbitals of the calculated fragment x and the ones from the embedding monomer y leading to near-orthogonality. The elements of the matrix \mathbf{P}^y take the following form in the AO basis functions χ_μ and χ_ν of the calculated fragment x :

$$P_{\mu\nu}^y = \langle \chi_\mu | \hat{p}^y | \chi_\nu \rangle. \quad (3.3)$$

Substitution with equation (3.2) yields

$$P_{\mu\nu}^y = \lambda \langle \chi_\mu | \left(\sum_{i \in y}^{n_{\text{occ}}^y} |\phi_i\rangle \langle \phi_i| \right) | \chi_\nu \rangle. \quad (3.4)$$

Here, representation of the molecular orbitals ϕ_i with the AO basis functions χ_σ and χ_ρ of the embedding fragment y together with the corresponding electron density matrix elements $D_{\sigma\rho}^y$ leads to

$$\begin{aligned} P_{\mu\nu}^y &= \lambda \langle \chi_\mu | \left(\sum_{\sigma=1}^{N_{\text{bas}}^y} \sum_{\rho=1}^{N_{\text{bas}}^y} |\chi_\sigma\rangle D_{\sigma\rho}^y \langle \chi_\rho| \right) | \chi_\nu \rangle \\ &= \lambda \sum_{\sigma=1}^{N_{\text{bas}}^y} \sum_{\rho=1}^{N_{\text{bas}}^y} \langle \chi_\mu | \chi_\sigma \rangle D_{\sigma\rho}^y \langle \chi_\rho | \chi_\nu \rangle = \lambda \sum_{\sigma=1}^{N_{\text{bas}}^y} \sum_{\rho=1}^{N_{\text{bas}}^y} S_{\mu\sigma} D_{\sigma\rho}^y S_{\rho\nu}. \end{aligned} \quad (3.5)$$

In matrix notation this writes as

$$\mathbf{P}^y = \lambda \mathbf{S} \mathbf{D}^y \mathbf{S} . \quad (3.6)$$

Thus, the shift matrix of an embedding fragment y can be constructed from the corresponding electron density matrix \mathbf{D}^y and the matrix \mathbf{S} which contains the overlap between basis functions of the calculated and the embedding fragments. Use of this relation is made within the course of the current chapter throughout.

Due to changes in the one-electron operator, the HF procedure of the fragment x is performed under the external Fock potential of all embedding monomers y . The application of shift operators ensures that the final molecular orbitals of x will result in a set that is near-orthogonal to all occupied embedding orbitals. Thus, the embedded calculation includes the same contributions involved in a complete HF treatment of all fragments. In the limit of exact Fock potentials the treatment of fragment x will not differ from the description of the same fragment within the full system HF calculation.

3.2 FJK

Parts of this section are featured in

"FJK - A cap-free fragment approach with embedding Fock potentials",
M. Werner and R. A. Mata, *AIP Conf. Proc.*, 1702, 09006, 2015.

For the application of Fock embedding within covalently bonded systems, a fragment approach to HF calculations was proposed, termed *Fragmentation with Coulomb (J) and exchange (K) embedding*, short FJK.^[66] This approach provides a method without parametrization of the boundary between fragments and allows for the convergence of fragment calculations when the embedding potential is correct in the limit of self-consistency. In this section the FJK method and its application for fragment HF calculations is investigated in detail. After the description of the method itself including the type of molecular fragmentation and computational details, the procedure is analyzed in terms of the resulting density and energy compared to the full HF solution of molecular systems. An approach making use of different basis sets representing the fragments and the potentials is investigated in the last part of the section.

3.2.1 Method

In order to employ the embedding with full Fock potentials for fragments of molecular systems, it is necessary to break covalent bonds. Since one is interested in fragments that may represent their state within the original environment, the generation of closed-shell monomers without introduction

of an internal error by the use of saturating link-atoms or molecular caps is desired. To achieve this aim, the simplest approach is the heterolytical fission of covalent bonds assigning both electrons of a covalent bond to one resulting fragment and none to the other. While a special proton shift procedure is applied in the FMO approach to avoid charged fragments^[58], the created FJK fragments may or may not carry a charge. This raises no particular issue as the counter-charge is covered by the embedding full Fock potential. Application of this fragmentation approach in form of shifting all binding electrons along the same direction within a long-chained system yields fragments of opposite charge only at the very end positions. The remaining monomers are electronically neutral. This is shown in Figure 3.1 for the example of propane.

From the generated fragments the FJK method approximates the full HF solution in a procedure consisting of four steps. The first three of these steps cover different type of monomer cycles where all fragments are calculated one after the other. Construction of the full system energy is carried out from the embedded monomer results in the last step. This procedure is displayed in Scheme 3.1.

In the first step 1) free fragment calculations are performed *in vacuo* to extract the first monomer Fock potentials. Within the second step 2) the fragments are calculated again embedded in the potentials of the other monomers. After this the potential of the calculated monomer is updated and directly used to embed the following fragment within the cycle. A fragment x in this monomer cycle is therefore embedded in the updated potentials of the fragments $y < x$ which have been already treated in this cycle, while the potentials of the remaining $y > x$ fragments are still the ones that have not been updated to this point. This cycle in the second step 2) can be repeated X times to further iterate the fragment potentials towards the full HF limit. The method is then referred to as FJK- X and each monomer cycle is denoted as a macro-iteration, while the term of a micro-iteration is used for single SCF steps within a fragment calculation.

With the second step 2) of the procedure an imbalanced description of the monomers is introduced. The last fragment of the corresponding monomer cycle is embedded in all the updated potentials, while only non-updated potentials are used to embed the first one. To overcome this

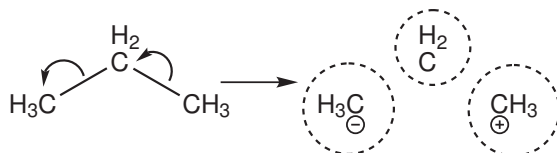


Figure 3.1: Heterolytical fragmentation in the FJK method on the example of propane.

Scheme 3.1: Procedure of the FJK method carried out in four steps. Different cycles over the fragments are carried out by the first three steps, while the energy of the total system is recovered in the last one.

1) Obtain starting potentials from free monomers:

$$\mathbf{h}_0^x \longrightarrow \mathbf{J}^x, \mathbf{K}^x, \mathbf{V}_n^x, \mathbf{D}^x$$

2) Update and direct use of monomer potentials:

$$\mathbf{h}^x = \mathbf{h}_0^x + \sum_{y=1}^{x-1} [2\tilde{\mathbf{J}}^y - \tilde{\mathbf{K}}^y + \tilde{\mathbf{V}}_n^y + \tilde{\mathbf{P}}^y] + \sum_{y=x+1}^N [2\mathbf{J}^y - \mathbf{K}^y + \mathbf{V}_n^y + \mathbf{P}^y]$$

$$\mathbf{h}^x \longrightarrow \tilde{\mathbf{J}}^x, \tilde{\mathbf{K}}^x, \tilde{\mathbf{V}}_n^x, \tilde{\mathbf{D}}^x \text{ repeat step 2) } X \text{ times} \Rightarrow \text{FJK-}X$$

3) Closing monomer cycle without any updates:

$$\tilde{\mathbf{h}}^x = \mathbf{h}_0^x + \sum_{y \neq x}^N [2\tilde{\mathbf{J}}^y - \tilde{\mathbf{K}}^y + \tilde{\mathbf{V}}_n^y + \tilde{\mathbf{P}}^y]$$

$$\tilde{\mathbf{h}}^x \longrightarrow \tilde{\mathbf{D}}^x, E^x$$

4) Energy calculation:

$$E_{\text{FJK}} = \sum_x^N E^x + \sum_{x=1}^N \sum_{y=x+1}^N \left\{ V_n^{x \leftrightarrow y} - \text{tr} [\tilde{\mathbf{D}}^x (\tilde{\mathbf{J}}^y - 1/2\tilde{\mathbf{K}}^y + \tilde{\mathbf{P}}^y)] \right\}$$

uneven fragment treatment another monomer cycle is carried out in step 3) of the procedure. This closing monomer cycle uses the latest potentials to embed all fragments without carrying out further updates of the potentials anymore. This ensures that all fragments are embedded with potentials of a consistent level balancing the resulting monomer description.

The fourth and last step 4) constructs the full system energy out of the embedded monomer calculations. In the first sum the energies of the embedded fragments are summed up, while nuclear repulsion between the fragments is added in the first term of the double sum. All electronic interactions between two fragments are calculated twice since a fragment x is embedded in the potential of another fragment y and vice versa. These double-counted interactions and the contributions of the non-physical shift operators are canceled out by the last term shown in step 4).

In contrast to other fragmentation approaches FJK does not introduce errors due to the introduction of hydrogen-link atoms or molecular caps for bond saturation. It also defines the properties of the total system solely on the basis of embedded monomer calculations without additional computation of dimers or trimers as commonly done in many-body fragment approaches like the FMO method of Kitaura *et al.*^[57], described in detail in section 2.2.2. While FMO performs embedded monomer calculations in the

Coulomb bath of the environment until self-consistency is reached, the full Fock potential of the surrounding fragments is included in FJK. A similar procedure with the additional inclusion of exchange potentials in monomer embedding of the FMO method has been tried, but to little success due to an imbalanced description of the consecutive 2-body calculations in the resulting field.^[61] The FMO method was only proposed to calculate the total energy of the system by a 2-body approach in a converged electrostatic field. In contrast to this, the characteristics of the FJK method are inspired by the mean-field SCF procedure itself and allow for the convergence towards the complete HF solution with an increasing number of macro-iteration monomer cycles. This property is analyzed in terms of the resulting energy and electronic density in the following section.

All calculations of this chapter have been carried out at the density-fitted Hartree-Fock (DF-HF) level of theory^[37] in combination with the cc-pVTZ basis set and corresponding fitting basis.^[67;68] A local development version of the MOLPRO2012.1 quantum chemistry package was used throughout.^[69] FJK results are always compared to the ones obtained by the full HF treatment. The parameter in the shift operators was set to $30000 E_h$ for all FJK fragment calculations. In order to avoid basis set superposition effects, each fragment FJK calculation has been carried out with the full molecular basis set. A mixed basis approach representing the embedding potentials in a minimal basis set is introduced in section 3.2.4.

3.2.2 Convergence Analysis

A simple case to check for the convergence of the proposed FJK method towards the full HF solution is given in the application to non-covalent molecular clusters. In these systems the break of chemical bonds is not required, since the fragments are naturally given by the monomers involved in the structure. For two different conformers of tetrameric water clusters taken from Mata and Stoll^[65], the convergence with an increasing number of macro-iterations was analyzed. These structures provide small neutral water molecules in a ring and a cage conformer allowing for the analysis of the FJK convergence without the effect of charged fragments generated by heterolytic bond fission. The FJK energy was calculated for both structures using the 4-step procedure as described above. Several repetitions of the uneven iterated monomer cycle were performed and the resulting energies were compared to the corresponding full system HF result. The structures and the differences in the absolute energies $\Delta E_{\text{abs}} = E(\text{FJK}) - E(\text{HF})$ are shown in Figure 3.2.

The results exhibit a fast convergence to the the full HF solution with errors below $2 \text{ kJ}\cdot\text{mol}^{-1}$ already after one repetition of the uneven monomer cycle in FJK-1. However, for the two water clusters, energy differences of opposite sign are found in the initial FJK-0 procedure. It was shown before

that the HF binding energies of these water tetramers differ with the ring conformer being the more stable one compared to the cage structure.^[65] Thus, the FJK-0 approach underestimates the binding of the ring conformer yielding a higher absolute energy, while the monomer interaction is overestimated for the cage conformer. For both structures this error vanishes with an increasing number of monomer cycles in the FJK procedure. A smooth convergence of the absolute energy compared to the full HF results is observed.

Another suitable criteria for the analysis of the convergence of the method is the resulting molecular electronic density. Since the electronic density of each fragment is repeatedly optimized in the potential of the surrounding monomers, the sum of all fragment densities should recover the full HF electronic density in the limit of self-consistency. The quantity of interest is therefore the difference $\Delta\rho(\mathbf{r}) = \rho_{\text{FJK}}(\mathbf{r}) - \rho_{\text{HF}}(\mathbf{r})$ between the electronic density observed by the FJK procedure and the full HF calculation. While the volume integral of this density difference will vanish due to the addition of positive and negative values, the integral of its absolute can be used to define another convenient quantity for the convergence analysis of the FJK method. Half of the value for the volume integral over the absolute density difference gives the quantity $\langle\delta n\rangle$ as

$$\langle\delta n\rangle = \frac{1}{2} \int |\Delta\rho(\mathbf{r})| \, \text{d}\mathbf{r} . \quad (3.7)$$

The value of $\langle\delta n\rangle$ can be taken as the mean number of displaced electrons

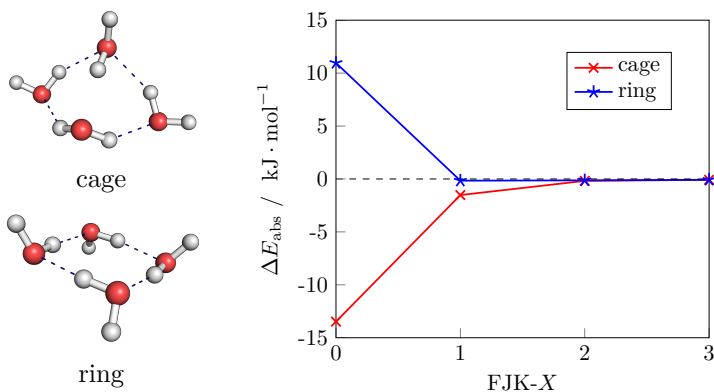


Figure 3.2: Left: Structure of two water tetramers in cage and ring conformation. Right: Convergence of the FJK energy compared to the full HF calculation $\Delta E_{\text{abs}} = E(\text{FJK}) - E(\text{HF})$ with an increasing number of macro-iterations for the two water tetramers.

described by the FJK method relative to the HF reference.

Both of the aforementioned quantities were applied to analyze the electronic density convergence of the FJK procedure in the propane system. As already seen in Figure 3.1, the molecule was divided into the three fragments $(\text{H}_3\text{C})^- \cdots (\text{CH}_2) \cdots (\text{CH}_3)^+$ and the resulting FJK density was constructed as the sum of all three monomer densities. The convergence due to FJK cycles is followed by isodensity surface plots of the density difference after each fragment calculation. In order to also analyze the effect of micro-iterations within the single fragments, the mean number of displaced electrons $\langle \delta n \rangle$ was calculated after each SCF step using a numerical grid of evenly spaced points with a step size of 0.183 \AA . All results of the FJK density convergence in the propane system are displayed in Figure 3.3.

Starting with an electronic density from free monomer calculations leads to a large mean number of displaced electrons $\langle \delta n \rangle$ of 1.57 compared to the full system HF result. This is also shown in the first surface plot where the electronic density on the left negatively charged methyl fragment is widely overestimated (blue contour), while the one from the right positively charged methyl fragment is too low compared to the HF density (red

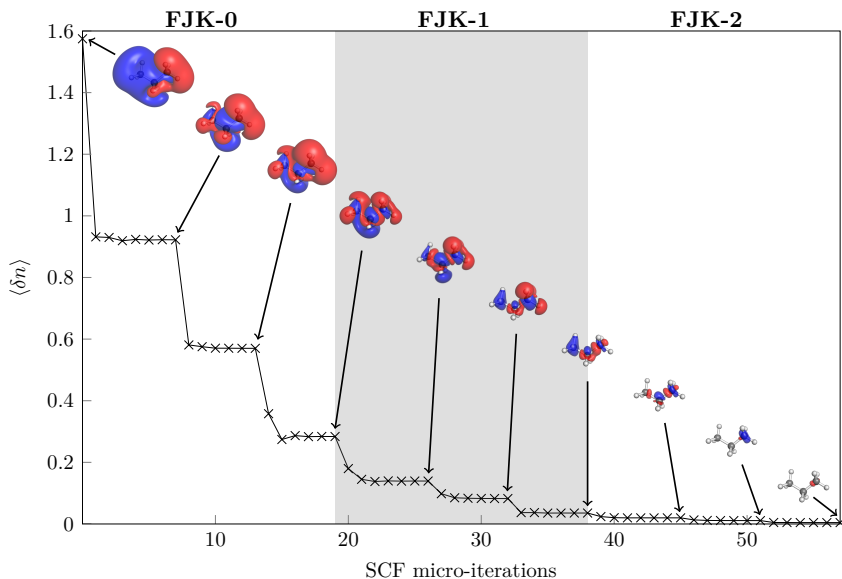


Figure 3.3: Development of the mean number of displaced electrons $\langle \delta n \rangle$ with each SCF step of the FJK method for propane. The graphics inside show isodensity surface plots of the density difference $\Delta\rho(\mathbf{r}) = \rho_{\text{FJK}}(\mathbf{r}) - \rho_{\text{HF}}(\mathbf{r})$ with a value of $5 \cdot 10^{-4} \text{ a}_0^{-3}$. Positive areas are colored in blue, negative ones in red. The FJK density was constructed as the sum of the embedded monomer energies.

contour). During the FJK macro-iterations the electronic densities of the monomers are improved from the $(\text{H}_3\text{C})^-$ over the (CH_2) to the $(\text{CH}_3)^+$ fragment. Already the first macro-iteration decreases the $\langle\delta n\rangle$ value to 0.28 at the FJK-0 level. Another monomer cycle in FJK-1 lowers the mean number of displaced electrons to 0.03, while the FJK density is basically converged at the FJK-2 level. This convergence is also seen in the surface plots, in which the density difference is further reduced with each monomer calculation. At the FJK-2 level only a very small contour indicates the remaining density difference.

Another interesting observation regarding Figure 3.3 is the fact that the main influence of the embedding potentials is already covered after a few micro-iterations in each fragment calculation. While the monomer density description is significantly changing during the first SCF steps in the corresponding fragment, its changes seem negligible when carrying out the final micro-iterations. This indicates the possibility to restrict the number of micro-iterations allowed in each fragment calculation without great loss of information on the density described. This option is further tested in the following conformational energy analysis of the FJK method.

3.2.3 Conformational Energy

Given that the FJK energy and density converge smoothly to the full HF result, proceeding to model applications seems feasible. In chemical processes one is usually not interested in the absolute energy of a system, but rather the changes of the energy due to occurring reactions, conformational changes or electronic excitations. This gives a more significant weight to the description of relative energies, which are analyzed for system of different size and complexity in this section.

In order to investigate the conformational energy description of FJK, a first benchmark included the alkane chains propane, butane, pentane and hexane. For all structures 36 conformers were generated by rotation around an internal C-C bond by 360° in steps of 10° . This corresponds to rotational barriers reaching from $15 \text{ kJ} \cdot \text{mol}^{-1}$ to $36 \text{ kJ} \cdot \text{mol}^{-1}$ at the HF level. The number of generated fragments is consistent with the number of carbon atoms n_C within the system yielding fragments as $(\text{H}_3\text{C})^- \cdots n_C \times (\text{CH}_2) \cdots (\text{CH}_3)^+$. Each conformer was calculated with the FJK method using different numbers of macro-iterations and with the full HF treatment. The three energy parameters used for conformational analysis include the mean absolute energy difference $\langle|\Delta E_{\text{abs}}|\rangle$, the maximum deviation of the relative energy MAXD_{rel} , and the relative root-mean-square deviation RMSD_{rel} . The results for all alkane systems are displayed in Table 3.1.

As it was already shown for the simpler cases in the previous section, the differences in absolute energies compared to the full HF results decrease with

Table 3.1: Mean absolute errors ($\langle |\Delta E_{\text{abs}}| \rangle$), relative maximum deviation (MAXD_{rel}) and relative root-mean-square deviation (RMSD_{rel}) in $\text{kJ} \cdot \text{mol}^{-1}$ of different FJK macro-iterations for alkanes from propane to hexane. The number of fragments was increased according to the number of carbon atoms involved as given in parenthesis behind the system.

Method	Propane (3 fragments)			Butane (4 fragments)		
	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}
FJK-2	9.3	0.05	0.03	30.8	2.24	1.67
FJK-3	1.1	< 0.01	< 0.01	4.6	0.81	0.51
FJK-4	0.4	< 0.01	< 0.01	1.0	0.18	0.10
Method	Pentane (5 fragments)			Hexane (6 fragments)		
	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}
FJK-2	54.1	0.65	0.25	82.5	1.50	0.69
FJK-3	8.8	0.57	0.33	15.3	1.12	0.65
FJK-4	1.6	0.14	0.08	2.7	0.29	0.18

a larger number of macro-iterations. At the FJK-4 level the mean absolute energy difference is already below $3 \text{ kJ} \cdot \text{mol}^{-1}$ for all alkane systems. A fast decrease is also observed for the maximum and root-mean-squared errors of the relative energies going from FJK-2 to FJK-4 indicating the robustness of the method. Using FJK-4 the procedure is capable of reproducing conformational HF energies with mean errors of less than $0.2 \text{ kJ} \cdot \text{mol}^{-1}$ even for the hexane system. Within this benchmark set also the trend concerning a larger number of generated fragments can be followed. Even though the errors in terms of absolute and relative energies are clearly rising for small numbers of macro-iterations from propane to hexane, the differences become less pronounced for higher FJK- X methods. The larger number of only roughly described fragments yields inferior starting points for the iterative procedure, but nevertheless the corresponding FJK energy can be converged to the full HF limit using slightly more updating cycles for the fragment potentials.

Two further benchmark systems, namely benzophenone and a capped dipeptide of alanine, were analyzed with two purposes in mind. On the one hand the generation of more exotic or even non-chemical fragments was investigated, on the other hand the idea of using a limited number of micro-iterations in each fragment calculation was pursued. The N-term of the alanine dipeptide was acetylated, while an N-methyl group was used to cap the C-term in order to avoid a charged system. For both systems 36 conformers were generated by 360° rotation around a C-C single bond in steps of 10° . This corresponds to HF rotational energy barriers of

$80 \text{ kJ} \cdot \text{mol}^{-1}$ for benzophenone and $65 \text{ kJ} \cdot \text{mol}^{-1}$ for the alanine dipeptide. The structures of both benchmark systems with the rotated single bonds highlighted in green are shown in Figure 3.4.

Both systems were divided into three fragments yielding a carbonyl group and two charged phenyl fragments in case of benzophenone, whereas the peptide unit CONH represents a single fragment of the alanine dipeptide system. The relative energy analysis was based on the same three parameters $\langle |\Delta E_{\text{abs}}| \rangle$, MAXD_{rel} and RMSD_{rel} , as already seen for the alkane benchmark set. While the FJK method with different numbers of macro-iterations was once applied with fully converged SCF cycles within the fragments, the procedure was repeated with the maximum number micro-iterations $n_{\text{maxit}}^{\text{micro}}$ per fragment being limited to four only. The corresponding results for both systems are displayed in Table 3.2.

It can be seen that the choice of more extreme monomers, like the charged phenyl fragments, goes hand in hand with larger errors in the absolute and relative energy parameters for FJK methods of lower order. Since these fragments are far from their original state within the molecule, the starting point for the iterative FJK procedure represents a poor description of the real molecular constitution. Nevertheless, this does not seem to raise particular issues as the absolute and relative energies still converge to the full HF results within a limited number of macro-iterations. For fully converged SCF cycles at the FJK-4 level, both systems reach a remarkable accuracy with average errors of the conformational energy within $0.05 \text{ kJ} \cdot \text{mol}^{-1}$.

With a restriction of the maximum number of micro-iterations per fragment calculation to four the overall energy observations are not changed drastically. This indicates that a few SCF steps per monomer are sufficient to cover the effect of the embedding potentials. The mean absolute energies for the restricted monomer treatment preserve the same trend as in case

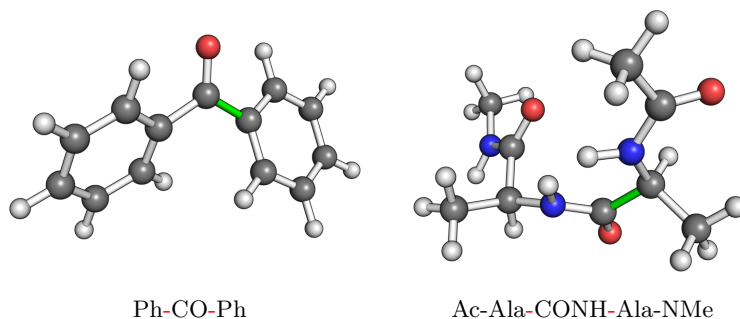


Figure 3.4: Structures of benzophenone (left) and the capped alanine dipeptide (right). The bonds around which the structures were twisted for generation of the different conformers are displayed in green. Both systems were split up into three fragments by cutting the bonds indicated with a red dash in the chemical formula below the structures.

Table 3.2: Mean absolute errors ($\langle |\Delta E_{\text{abs}}| \rangle$), relative maximum deviation (MAXD_{rel}) and relative root-mean-square deviation (RMSD_{rel}) in $\text{kJ} \cdot \text{mol}^{-1}$ of different FJK macro-iterations for benzophenone and the alanine dipeptide. The upper part includes full SCF cycles while the maximum number of micro-iterations per fragment was limited to four in the lower part of the table.

Method	Benzophenone			Alanine dipeptide		
	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}
full SCFs						
FJK-2	34.2	2.19	1.25	36.3	2.09	1.14
FJK-3	4.3	0.53	0.33	5.4	0.45	0.17
FJK-4	0.8	0.10	0.04	1.3	0.13	0.05
$n_{\text{maxit}}^{\text{micro}} = 4$						
FJK-2	53.5	1.14	0.66	50.1	2.79	1.60
FJK-3	6.1	0.40	0.25	8.5	0.33	0.17
FJK-4	1.8	0.50	0.23	2.0	0.29	0.22

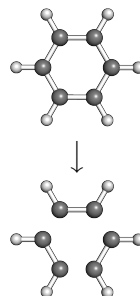
of the converged fragment calculations. With slightly larger deviations than for the fully converged fragments, the absolute energies described by restricted monomers are still converging to the full system HF result with an increasing number of macro-iterations. The same is not true for the behavior of the relative energy. In fact, for the example of benzophenone at the FJK-2 and FJK-3 level, the errors in the relative energy quantities are even smaller than the ones for fully converged monomer SCF cycles. The errors caused by the non-converged description of the fragments are able to add up or compensate in a different way for each conformer. Only with a sufficient number of macro-iterations minimizing the error in each fragment description one will observe a stable behavior in the relative energy.

One of the possible limitations of the FJK method is the arbitrary choice of fragments allowing for the generation of non-physical fragments. It is uncertain how this fragmentation may impact on the resulting energy description and the asymptotic convergence to the full HF solution. The effect of a critical fragment choice was investigated for the extreme case of a completely delocalized system. A single molecule of benzene was split into three fragments of C_2H_2 . 14 conformers were generated for this system by varying the C-C bond distance from 1.32 Å to 1.45 Å in steps of 0.01 Å. This corresponds to a total HF conformational energy window of $30 \text{ kJ} \cdot \text{mol}^{-1}$. The parameters of the absolute and relative energy differences for a variety of FJK macro-iterations and the structure of benzene together with its fragmentation are shown in Table 3.3.

It is easily seen that a larger number of FJK macro-iterations is needed

Table 3.3: Mean absolute errors ($\langle |\Delta E_{\text{abs}}| \rangle$), relative maximum deviation (MAXD_{rel}) and relative root-mean-square deviation (RMSD_{rel}) in $\text{kJ} \cdot \text{mol}^{-1}$ of different FJK macro-iterations for benzene. The C_6H_6 molecule is split into three neutral monomers of C_2H_2 .

Method	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}
FJK-2	122.3	20.36	12.50
FJK-3	45.6	10.31	2.68
FJK-4	16.1	4.71	2.68
FJK-5	5.6	1.82	0.96
FJK-6	2.5	0.42	0.27
FJK-7	1.5	0.20	0.10



to achieve the same accuracy in the benzene molecule as in the previous cases. Starting from large deviations in the lower order FJK methods, mean relative errors below $1 \text{ kJ} \cdot \text{mol}^{-1}$ are found from FJK-5 on. The calculation is basically converged at the FJK-7 level with an mean error in the absolute energies of only $1.5 \text{ kJ} \cdot \text{mol}^{-1}$. This is a clear indication that the asymptotic convergence of the FJK procedure is independent of the fragment choice. Even charged fragments and monomers far from their natural chemical environment can be used to recover the full HF result in the limit of the self-consistent description of their potentials. Also in the case of delocalized systems the complete HF limit is reached with an increasing number of macro-iterations, which allow the monomer densities to spread among each other.

This section solely investigates the convergence of the FJK method to the full HF limit with an increasing number of macro-iterations, while no focus is set to the corresponding computational cost. In its current form the FJK method is carried out in the full molecular basis set for all fragment calculations. Even though this allows for the analysis of the resulting errors free from basis set superposition effects, it does not provide the full benefits known from other fragmentation approaches. The use of all basis functions in every fragment treatment leads to an unfavorable scaling of the FJK procedure with growing system size. Even with a limited number of micro-iterations the FJK method is computational more costly than the full HF calculation. The fragment calculations involved do not provide pure monomer treatments, since their computational effort is not independent of the total systems size due to the use of basis functions on the embedding fragments. However, in order to represent the embedding Fock potentials basis functions on the surrounding fragments are mandatory. The included Fock potentials are based solely on the corresponding fragment density. This points out the possibility to use a minimal basis set for the description of the embedding fragments providing a rough representation

of the corresponding monomer densities. Whether this idea is sufficient to include the embedding Fock potentials, is analyzed in detail in the following section.

3.2.4 Mixed Basis Approach

The FJK procedure was modified by the use of a minimal basis set for the embedding fragments, while the larger basis remains only on the calculated monomer. This mixed basis FJK (MBFJK) description introduces several additional operations within the aforementioned FJK steps, since the monomers are now treated with different basis sets and their potentials need to be transferred into the basis used for their embedding. Any electronic density in the AO basis of one fragment y can be projected to the AO basis of the fragment x by using the overlap matrix between both fragments \mathbf{S}^{xy} with the corresponding elements $S_{\mu\sigma}^{xy} = \langle \chi_{\mu}^x | \chi_{\sigma}^y \rangle$. This projection writes as

$$\mathbf{D}^x = \mathbf{S}^{xy} \mathbf{D}^y \mathbf{S}^{yx} . \quad (3.8)$$

With the projection of fragment densities it is possible to build the embedding potential of a surrounding monomer in the basis of the calculated entity. A schematic procedure of the MBFJK approach and the involved operations are given in Table 3.4.

As in the full basis FJK approach all fragments are first calculated individually to obtain their initial densities. In the case of the MBFJK method this treatment occurs in a minimal basis set representation of the calculated monomers. This eases the projection of the fragment densities in the following step, as the minimal basis description is a subset of the basis set used to include the fragment as an embedding unit. All embedding densities are projected into to the mixed basis of the treated fragment unit x , which describes x by a larger basis set, while the remaining fragments are represented in a minimal set of AO functions. These projected densities are then used to build the embedding potentials which are included in the one-electron operator of x . The update of the monomer densities and thus the embedding potentials if performed in an asymmetric cycle as in the full basis FJK approach and can be repeated by several macro-iterations to improve the fragment description.

A clear contrast to the full basis FJK procedure is the calculation of the total energy in the MBFJK method. Since the various fragments are treated with different basis sets, an energy composition out of their individual fragment energies leads to an unbalanced description of the system. With the projected and updated property for each fragment now being the corresponding electronic density, this is also the quantity that needs to be applied to achieve a stable and converging treatment of the total system. The final electronic densities of all fragments are therefore

Table 3.4: Schematic procedure for the MBFJK approach in a system of N fragments. The table includes information on the operations carried out for the given fragments and the corresponding basis set used.

Fragments	Operations	Basis
$\forall x \in \{1\dots N\}$	micro SCF cycle $\rightarrow \mathbf{D}^x$	x @ minimal
macro-iterations $\left\{ \begin{array}{l} \forall x \in \{1\dots N\} \\ \forall y \neq x \end{array} \right.$	project \mathbf{D}^y , form $\mathbf{J}^y, \mathbf{K}^y$,	
	load \mathbf{V}_n^y and form \mathbf{P}^y	x @ large
	$\mathbf{h}_0^x + 2\mathbf{J}^y - \mathbf{K}^y + \mathbf{V}_n^y + \mathbf{P}^y$	rest @ minimal
	micro SCF cycle $\rightarrow \tilde{\mathbf{D}}^x$	
$\forall x \in \{1\dots N\}$	project $\tilde{\mathbf{D}}^x$, $\mathbf{D}_{\text{tot}} = \sum_{x=1}^N \tilde{\mathbf{D}}^x$	all @ large
full geometry	one full HF step	all @ large

projected into the full molecular basis set with functions of the larger basis placed on all atoms. Summing up the resulting projected densities to a total electronic density yields a suitable starting point for the calculation of the MBFJK energy. A Fock matrix is constructed from the obtained electronic density and subsequently diagonalized in a final conventional full HF step. This ensures a balanced energy and density description of the system and makes the approach accessible to all conventional post-HF methods.

The MBFJK method was applied to calculate the conformational energies of four of the benchmark systems already used to analyze the FJK performance, butane, hexane, benzophenone and benzene. The large basis set used for fragment calculation was cc-pVTZ^[67], while STO-3G^[70;71] was used as a minimal basis set for potential representation. An auxiliary set of functions for density fitting was only placed on the calculated fragment.^[68] The established parameters for analysis of absolute and relative energy description compared to the full HF calculations were used and can be found for different macro-iterations of the MBFJK approach in Table 3.5.

It becomes clear that the MBFJK method yields absolute and relative energies close to the full HF results already with a small number of macro-iterations carried out. Due to the final full HF step from embedded fragment densities the mean errors in the relative energy are below $2 \text{ kJ} \cdot \text{mol}^{-1}$ for all four systems already at the MBFJK-0 level. The effect of the final full HF step is quite pronounced, which can be seen by comparison of the results for hexane with the ones from the full basis FJK

Table 3.5: Mean absolute errors ($\langle |\Delta E_{\text{abs}}| \rangle$), relative maximum deviation (MAXD_{rel}) and relative root-mean-square deviation (RMSD_{rel}) in $\text{kJ} \cdot \text{mol}^{-1}$ of different MBFJK macro-iterations for butane, hexane, benzophenone and benzene. The large basis set used was cc-pVTZ, the minimal basis set STO-3G.

Method	Butane (4 fragments)			Hexane (6 fragments)		
	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}
MBFJK-0	5.4	0.02	0.01	14.8	1.24	0.72
MBFJK-1	3.5	0.02	0.01	6.0	0.18	0.10
MBFJK-2	3.4	0.02	0.01	5.6	0.08	0.05

Method	Benzophenone (3 fragments)			Benzene (3 fragments)		
	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}	$\langle \Delta E_{\text{abs}} \rangle$	MAXD_{rel}	RMSD_{rel}
MBFJK-0	9.8	0.54	0.35	29.4	2.42	1.52
MBFJK-1	9.4	0.51	0.33	11.5	1.32	0.80
MBFJK-2	9.6	0.53	0.33	8.5	0.85	0.50

procedure in Table 3.1. Several macro-iteration cycles are needed for a stable description of the absolute and relative energies from embedded monomer calculations in the initial FJK method. The MBFJK approach benefits from the combination of monomer densities in a final HF step leading to errors comparable to the conventional FJK-4 results already at the MBFJK-1 level.

By application of the MBFJK procedure to both alkane systems, butane and hexane, the energy parameters are quickly found to be close to the full HF limit with an increasing number of macro-iterations, the convergence is significantly slower in the cases of benzophenone and benzene. In the latter systems the MBFJK procedure is incapable of smoothly recovering the conjugation of the structures. As the provided orbital space on the embedding fragments is drastically reduced, the monomer density is less effectively spread among the other fragments. However, for the description of relative conformation energies the MBFJK-2 approach yields mean differences to the full HF results of less than $1 \text{ kJ} \cdot \text{mol}^{-1}$ for all four benchmark systems.

Considering the fragment operations involved in the MBFJK procedure and the formal scaling of the HF method, a rough estimation for the development of the computational cost with a rising number of fragments can be estimated. This is easiest to follow in a system consisting of N_{frag} equal fragments, like in the case of a water cluster. Each monomer may be described at a larger basis set with N_{large} functions for its calculation or at a smaller basis set with N_{small} functions for its embedding. Assuming that a HF calculation scales cubically with the number of used basis functions,

the full calculation of the total system with a large AO basis yields an increase of the computational time t_{comp} as

$$t_{\text{comp}}(\text{HF}) \propto [N_{\text{frag}} \cdot N_{\text{large}}]^3 . \quad (3.9)$$

Considering only the free monomer and mixed basis fragment calculations in one macro-iteration cycle of MBFJK, the computational effort would increase with the term

$$t_{\text{comp}}(\text{MBFJK}) \propto N_{\text{frag}} \cdot [N_{\text{small}}]^3 + N_{\text{frag}} \cdot [(N_{\text{frag}} - 1) \cdot N_{\text{small}} + N_{\text{large}}]^3 . \quad (3.10)$$

The dependency of both computational costs on the number of fragments N_{frag} is shown in Figure 3.5 with $N_{\text{large}} = 58$ and $N_{\text{small}} = 7$, as it would be the case in a water cluster with a large cc-pVTZ and a small STO-3G basis set.

The graph exhibits a 'sweet spot' for the MBFJK method with a favorable scaling compared to the full HF calculation in an area between roughly 100 and 550 fragments. However, it also clearly shows that the computational cost of MBFJK is rising more steeply than the one of HF making the fragment method unfavorable for larger systems. In any case this estimation represents an ideal scaling of the MBFJK method. For simplicity, the proportionally factors to the computational time of MBFJK and HF were assumed to be equivalent. These prefactors will impact on the actual computational scaling and differ among both calculations with a strong dependence on the investigated system. In addition to this, only the fragment calculations of MBFJK were considered. Taking into account the additional operations of monomer density projections, generation of

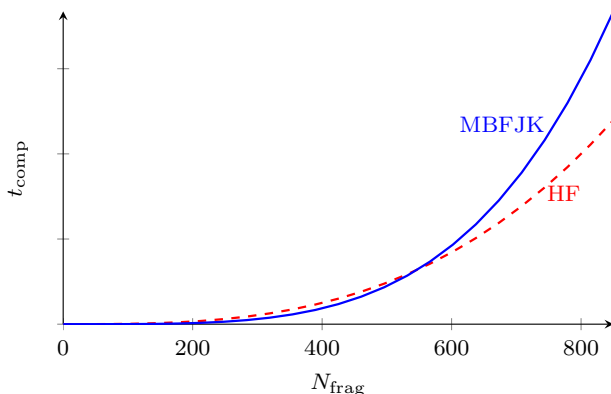


Figure 3.5: Comparison of the formal scaling of conventional Hartree-Fock and the MBFJK approach with an increasing system size. The number of basis functions per fragment corresponds to a water cluster with a large cc-pVTZ and a small STO-3G basis set.

embedding potentials, repetition of the macro-iteration monomer cycle and the final full HF step, the actual scaling will yield an inferior picture where a very narrow favorable area compared to the full HF calculation can be found at best.

This rough scaling estimation excludes the FJK procedure as a fragmentation method recovering the full HF solution at a general lower computational cost. Due to the fact that basis functions are needed to represent the embedding potentials and allow the fragment densities to spread over the neighboring monomers, it is difficult to achieve pure fragment calculations and the corresponding favorable computational scaling. Even with a mixed basis approach representing the embedding potentials in minimal basis sets or with a limited number of performed micro-iterations in the SCF cycle of each fragment, the resulting computational timing is inferior to the full HF description for all systems of chemical application. Thus, the full Fock embedding was not successful to yield a fragmentation approach with linear scaling with the system size as it would be the case for independent monomer calculations. However, the FJK method demonstrated that full Fock embedding can be used to describe monomer energies and densities under the effect of their neighbors in a limit that converges to the full HF treatment. This property is taken up in the following section, making use of full Fock potentials in the analysis of standard QM/MM embedding techniques.

3.3 Embedding Analysis

The use of quantum mechanics/molecular mechanics (QM/MM) hybrid methods^[72] is an approach of evolving success in the past decades for description of chemical phenomena within an embedding environment.^[73–75] Its applications reach from reactions in the active sites of enzymes over explicit solvent simulations up to periodic structure calculations in solid states.^[76;77] The interaction between the quantum mechanical and molecular mechanical regions is a key element to the success of these QM/MM calculations and can be included by different embedding schemes. At the simplest level the interaction is calculated at the MM level in *mechanical embedding* schemes. A more sophisticated approach that is nowadays commonly used is given in form of *electrostatic embedding* schemes, which make use of point charge representations of the MM area to embed the QM region allowing for its polarization. This polarization effect is mandatory to accurately describe chemical phenomena inside of the QM part of the system.^[78] It was recently shown, that also the repolarization of the MM region, as taken into account in *polarized embedding* schemes, can have a significant impact in the properties under consideration.^[79]

Polarization effects considered in QM/MM approaches are in most cases restricted to Coulomb contributions represented by embedding point

charges. Especially for the closest environment of the QM region this treatment is lacking the inclusion of other polarization contributions leading to artifacts within the resulting QM description. The possible overpolarization of the QM region by very close point charges is a problem often discussed in this context.^[80;81] To overcome these limitations embedding approaches based on fixed quantum mechanical descriptions of the environment were established, like the frozen orbital approach for wave function methods^[82;83] or the frozen density embedding formulation in DFT.^[84;85] Recently, also polarizable density embedding approaches have been proposed in order to accurately represent the electrostatic potential of the environment.^[86;87] With the purpose in mind to describe the embedding environment in a fashion close to the exact HF treatment, the use of full Fock potentials is applied to the analysis of standard QM/MM embedding approaches in the following section.

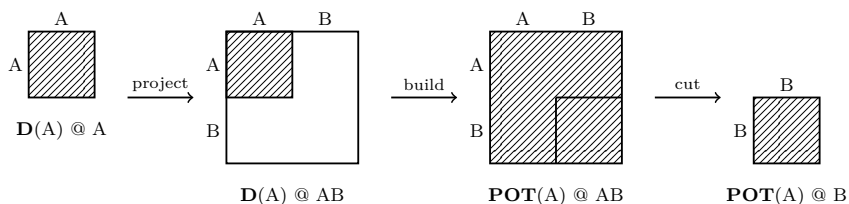
3.3.1 Procedure

For the comparison of the embedding with full Fock potentials to standard QM/MM embedding techniques, it is necessary to restrict the contributions of an embedding monomer to the basis set used in the QM region. This is achieved by a superposition approach combining the basis sets of the calculated and the embedding monomers. For the transfer of the embedding potential from one fragment to another, a schematic illustration of the approach is shown in Scheme 3.2.

In order to transfer the contributions of the potential of an embedding monomer A to the basis functions of a monomer B, both entities are first calculated individually in their own basis sets. Subsequently, the orbitals of A are explicitly orthogonalized against the orbitals of B to account for their initial orbital interaction. In a following step the electronic density of A is projected to the combined basis sets of A and B, in which it is only a submatrix with entries for the basis functions of A. The potential of A including the Coulomb, exchange, nuclear and shift operator contributions is then constructed out of the electronic density in this combined basis set. A final cut of the resulting matrix is performed extracting only the entries of the basis functions of B. This matrix is eventually used in the embedded calculation of monomer B by adding it to the corresponding one-electron operator. The superposition approach allows for the effect of Fock embedding solely in the basis of monomer B and thus for the direct comparison to standard QM/MM embedding schemes.

By only building the Coulomb but not the exchange contribution of the potential, the missing effect to the QM region compared to the full Fock embedding can be analyzed. Shift operators of all embedding monomers were used throughout as an external potential with $\lambda = 1000 E_h$ in order to make all embedding results comparable. Several point charge models

Scheme 3.2: Superposition approach for the transfer of the potential of a monomer A into the basis set of monomer B. The steps and modifications in terms of the involved electron density and Fock potential matrices are shown.



used in current QM/MM studies were compared to the embedding with full Fock potentials. The lacking effects in point charge approaches were analyzed for the first and second molecular shell around the QM region by comparison of the resulting monomer density. The example of an effect of higher-order was covered by the calculation of the dispersion energy of molecular dimers embedded in different potentials.

3.3.2 Monomer Density

For a simple test case, the water dimer, the density of both monomers were analyzed with different potentials used to represent the other monomer. As already seen for the FJK convergence of the electronic density in chapter 3.2.2, the mean number of displaced electrons $\langle \delta n \rangle$ as given in equation (3.7) was used as a parameter to quantify density differences to the full Fock embedding reference in $\Delta\rho(\mathbf{r}) = \rho_{\text{pot}}(\mathbf{r}) - \rho_{\text{Fock}}(\mathbf{r})$. Monomer densities of both water molecules were obtained for an embedding representation by

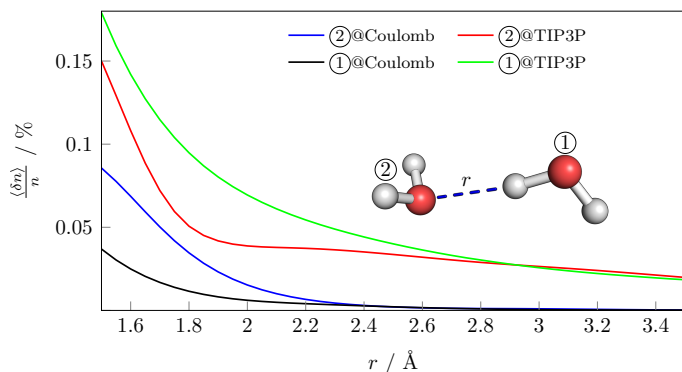


Figure 3.6: Percentage of mean displaced electrons of a water molecule with an approaching monomer represented by different potentials compared to full Fock embedding. The monomer in the legend is represented by the labeled potential, while the other one is calculated at the QM level.

TIP3P^[88] point charges and the Coulomb potential with missing exchange contributions for different dimer distances. The percentage of the resulting mean numbers of displaced electrons compared to full Fock embedding are shown in Figure 3.6.

It is seen that calculation of monomer ② in the TIP3P charges of monomer ① yields large mean numbers of displaced electrons even for distant dimers. The errors in the resulting electronic density increases when both monomers are approaching each other due to the poor description of strong intermolecular interactions. In the swapped situation of monomer ① in the TIP3P point charges of ② the error is close to constant in a broader area and only exhibits a steep increase for distances smaller than 1.8 Å. This difference is easily rationalized by looking at the closest point charge to the QM region. In the first case the charge is of positive sign located at the hydrogen atom of ① leading to a spread of the monomer density of ② towards the point charge. For the calculation of monomer ① the negative charge on the oxygen atom of ② is the closest, which yields a less significant polarization of the monomer density.

In both cases of Coulomb embedding the error compared to the full Fock potential is solely defined by the missing exchange contributions. Thus, the mean number of displaced electrons is vanishing for very distant monomers, as their exchange interaction rapidly decays. Only for very short distances a slightly more disordered monomer density is obtained due to the increasing effect of exchange contributions in the dimer interaction.

With this approach the full Fock embedding is suitable for the analysis of monomer densities within an external potential and directly comparable to different point charge models commonly applied in QM/MM embedding. Therefore, several model structures from current QM/MM studies were used to analyze the electronic density of single monomers embedded in different potentials. For three QM/MM snapshots from liquid and solid phase the effect of an embedding first monomer shell around the central unit was analyzed. Geometries of a CO₂ crystal, a cyanamide crystal and

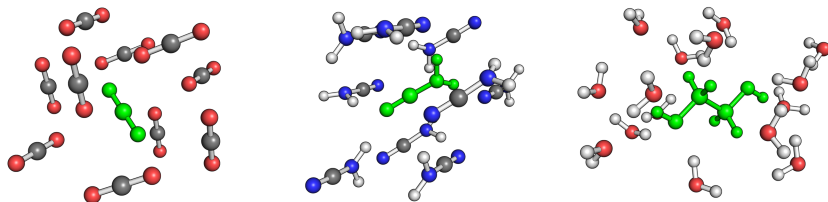
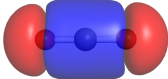
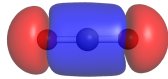
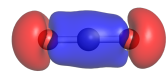
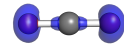
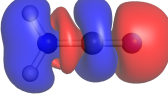
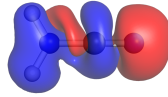
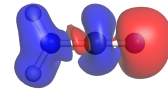
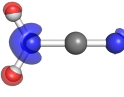
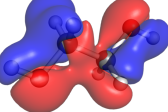
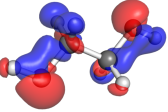
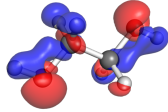
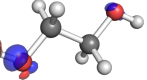


Figure 3.7: CO₂ crystal structure (left), cyanamide crystal structure (middle) and ethylene glycol in water (right) used for the monomer density analysis with different types of embedding. The calculated monomer embedded in the potentials of the others is shown in green.

Table 3.6: Density analysis for different embedding schemes of a monomer in the CO₂ crystal (top), the cyanamide crystal (middle) and ethylene glycol in water (bottom) compared to the full Fock embedding. Density differences $\Delta\rho(\mathbf{r}) = \rho_{\text{pot}}(\mathbf{r}) - \rho_{\text{Fock}}(\mathbf{r})$ are displayed with an isosurface value of $5 \cdot 10^{-4} \text{ a}_0^{-3}$. Positive values are colored in blue, negative ones in red. The percentage of mean displaced electrons compared to full Fock embedding and the corresponding difference in the correlation energy of a consecutive DF-MP2 calculation are provided.

	free	GAFF	iterated NPA charges	Coulomb
				
$\frac{\langle \delta n \rangle}{n} / \%$	0.25	0.13	0.04	0.01
$\Delta E_{\text{corr}} / \frac{\text{kJ}}{\text{mol}}$	-12.5	-6.6	-1.9	0.2
	free	GAFF	iterated NPA charges	Coulomb
				
$\frac{\langle \delta n \rangle}{n} / \%$	0.90	0.46	0.28	0.03
$\Delta E_{\text{corr}} / \frac{\text{kJ}}{\text{mol}}$	-20.6	-10.9	-7.0	0.3
	free	TIP3P	TIP3P+exchange	Coulomb
				
$\frac{\langle \delta n \rangle}{n} / \%$	0.44	0.12	0.11	0.01
$\Delta E_{\text{corr}} / \frac{\text{kJ}}{\text{mol}}$	-10.1	-3.1	-2.9	0.2

an ethylene glycol molecule in water were used to embed the central QM molecule in the potentials of the closest 10, 9 and 13 monomers, respectively. The corresponding structures are shown in Figure 3.7.

The monomer HF calculations in different embedding potentials were analyzed in terms of their electronic density and the corresponding correlation energy of a consecutive density fitted Møller-Plesset perturbation theory^[89] (DF-MP2). While no embedding potential was used as a first reference compared to the full Fock embedding, the QM/MM point charge models covered the fitted HF/6-31G* RESP charges^[90] as commonly employed by the generalized amber force field (GAFF)^[91] for the crystals and TIP3P^[88] charges for the water molecules. For both crystal structures point charges

obtained by natural population analysis (NPA)^[92] of an iterative QM/MM procedure similar to the one from Bjornsson and Bühl^[77] were used for embedding as well. Coulomb embedding was applied to all systems in order to quantify the effect of missing exchange contributions compared to the Fock potentials. The resulting density differences to the full Fock embedding are shown as isodensity surface plots together with their mean number of displaced electrons and the difference in correlation energy in Table 3.6.

Comparing the free central monomer calculation to the one with full Fock embedding leads to large contours in the isodensity surface plots. Blue contours indicate an overestimation of the monomer density compared to full Fock embedding, while a red contour represents a density deficit. The missing intermolecular interactions also have an impact on the correlation energy with differences of more than $10 \text{ kJ} \cdot \text{mol}^{-1}$ in all three systems.

Considering the percentage of mean displaced electrons, the total embedding effect of the surrounding monomers on the calculated electronic density can be estimated. The density changes due to the inclusion of full Fock potentials is largest for the strongly hydrogen bonded cyanamide crystal structure with 0.90% of all electrons being displaced. A smaller embedding impact on 0.44% of all electrons is observed for the ethylene glycol solute in water. Only slight density adjustments effecting 0.25% of the electrons are found in the case of the non-polar CO_2 crystal structure.

For both crystal structures the application of GAFF charges for monomer embedding covers roughly half of the effect compared to full Fock potentials in terms of mean number of displaced electrons and correlation energy. The TIP3P description of water molecules around the ethylene glycol solute yields an improved density of the central unit with an absolute error in the correlation energy of only $3.1 \text{ kJ} \cdot \text{mol}^{-1}$. While the use of iterated charges observed by NPA gives results close to the full Fock limit for the non-polar CO_2 crystal, the resulting density is not satisfying in the case of the strongly interacting cyanamide. The point charge model is not sufficient to describe the directed hydrogen bonds involved yielding a persisting absolute error in the correlation energy of $7.0 \text{ kJ} \cdot \text{mol}^{-1}$. In all three systems the use of full Coulomb potentials for monomer embedding leads to only spurious remaining contours in the isodensity surface plots and to no significant errors in the described monomer density. This is a clear indication that the Coulomb contributions of the Fock potentials are the dominating factor influencing the monomer density changes. The effect of the exchange contributions is by several orders of magnitude smaller for all systems considered.

So far, all QM/MM systems have been analyzed in terms of the embedding effect of a first molecular shell around the calculated central monomer. In order to also investigate the impact of a secondary monomer shell, the crystal structure of ammonia was split into two different layers around a

central monomer unit. The first shell around a central ammonia molecule included 12 monomers in a sphere with a radius of 4 Å. 24 monomers outside of the sphere were used to represent the secondary ammonia shell. The corresponding structure is shown in Figure 3.8.

Effects of both ammonia layers were investigated separately either by embedding with monomers of the first monomer shell represented by different potentials or by inclusion of monomer potentials from the second shell, while full Fock embedding was used to represent monomers in the first shell. The central molecule was calculated *in vacuo*, embedded in point charges from the GAFF, in charges obtained by NPA from an iterative QM/MM procedure and in the full Coulomb potential of the investigated ammonia shell. The errors compared to a full Fock potential of the layer under study were analyzed in terms of the mean number of displaced electrons and the consecutive DF-MP2 correlation energy. The results are shown in Table 3.7.

It is seen that for both criteria that the effect on the calculated monomer density is smaller for the second than for the first ammonia shell. All types of point charge and Coulomb embedding exhibit no significant errors in the mean number of displaced electrons and the correlation energy for the embedding potentials of the secondary ammonia shell, as long as the first shell is represented by full Fock embedding. For the first shell analysis similar trends to the already analyzed QM/MM crystal structures are observed. Due to the specific interactions of the ammonia molecules via hydrogen bonds, the effect of full Fock embedding is only partly covered by simple GAFF charges. The use of iterated NPA charges once again improves the description, but is still lacking part of the interaction effect compared to full Coulomb and Fock embedding. As already observed

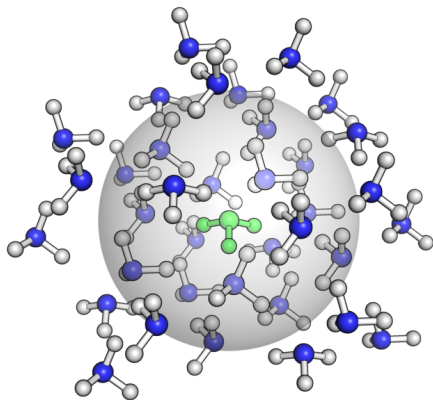


Figure 3.8: Snapshot of an ammonia crystal structure. The gray sphere indicates the first molecular shell around the calculated monomer shown in green. Molecules on the outside of the sphere are assigned to the secondary ammonia shell.

Table 3.7: Comparison of the mean displaced electrons and correlation energies from DF-MP2 for potentials used to represent the first and second molecular shell in an ammonia crystal. The results are referenced to the corresponding full Fock potential. For the analysis of the second shell the first shell was included via full Fock embedding.

free		GAFF		iterated NPA charges		Coulomb	
first	second	first	second	first	second	first	second
$\langle \delta n \rangle_n / \%$							
0.53	0.05	0.40	0.04	0.12	0.04	0.04	0.04
$\Delta E_{\text{corr}} / \text{kJ} \cdot \text{mol}^{-1}$							
-1.37	0.03	-1.10	0.02	-0.40	-0.03	-0.13	-0.06

before, the exchange contributions seem to have a minor impact on the resulting monomer density.

The QM/MM embedding analysis of this section focused on the changes in the calculated molecular density. This gives an estimate for the effect of different embedding techniques on first-order molecular properties which are directly observed from the resulting electronic density. An example for the effect of embedding potentials on a property of higher-order is given in the following section in form of the influence on the dispersion interaction energy of molecular dimers.

3.3.3 Dispersion energy

In order to analyze the impact of different embedding potentials on the dispersion interaction energy of molecular dimers, two QM/MM structures were chosen. Two monomers of a formamide crystal structure were embedded in the potentials of the 14 closest molecules, while a structure of a butanone solute in toluene was used to analyze the effect of 13 surrounding solvent molecules on a butanone toluene dimer. Two snapshots of both QM/MM structures are shown in Figure 3.9.

The dispersion interaction energies of both dimers were calculated within the full Fock potentials of the embedding molecules and compared to the free dimer calculation, the dimer in point charges and to full Coulomb embedding. The QM/MM point charges for the solid and liquid QM/MM snapshots were taken following the fitting procedure of the generalized amber force field (GAFF)^[91] and the optimized potentials for liquid simulations (OPLS)^[93], respectively. The embedded HF reference calculation was followed by a density fitted spin component scaled local MP2 calculation with a cc-pVTZ basis set (DF-SCS-LMP2/cc-pVTZ).^[1;94] To obtain the dispersion interaction energy E_{disp} a decomposition scheme of individual

Table 3.8: Dispersion interaction energies of a formamide and a butanone toluene dimer in full Fock embedding and its change using different potentials for the environment. E_{disp} was calculated using a decomposition scheme at the DF-SCS-LMP2/cc-pVTZ level. All values are provided in $\text{kJ} \cdot \text{mol}^{-1}$.

System	E_{disp}	ΔE_{disp}		
		free	GAFF/OPLS	Coulomb
formamide crystal	-3.03	-0.14	-0.03	< 0.01
butanone in toluene	-7.86	0.04	0.02	< 0.01

orbital excitation contributions as described by Wuttke and Mata was applied.^[95] The results of the dispersion energies for both systems in the full Fock potential and their changes due to other embedding schemes are displayed in Table 3.8.

It first can be seen that the total dispersion interaction energy of the butanone toluene dimer is more than two times larger than in the case of the formamide dimer. This is understood in terms of the less polar and larger monomers involved in the solute solvent dimer summing up more individual dispersion contributions to the total value of E_{disp} . However, no significant changes in the dispersion interaction energies using different potentials such as point charges or full Coulomb embedding are observed. Even comparing the full Fock embedded and the free formamide dimer lowers the corresponding dispersion interaction energy only by $0.14 \text{ kJ} \cdot \text{mol}^{-1}$, which is below the intrinsic error of the method applied.

The inclusion of an embedding potential for a molecular dimer calculation yields only marginal changes in the corresponding dispersion interaction energy. It can therefore be concluded that the interactions with the environment have a minor impact on the dispersion interaction of a molecular

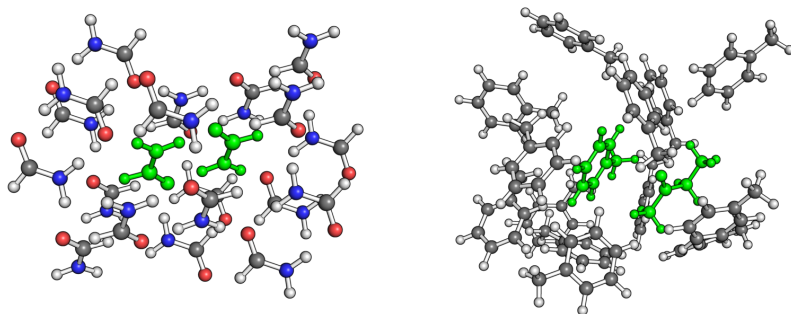


Figure 3.9: Structures of a formamide crystal (left) and butanone in toluene (right). Dimers used for the analysis of their dispersion energy in different embedding potentials are shown in green.

dimer, since the main contribution is determined by orbital excitations depending on the physics of the dimer itself.

3.4 Conclusion

The complete Fock potentials of surrounding fragments or monomers can be used for embedding of a Hartree-Fock calculation by changing the effective one-electron operator. Two different applications of this type of embedding were analyzed. On the one hand the use of full Fock embedding for fragment calculations can ensure the convergence to the full Hartree-Fock description in an iterative limit, on the other hand the embedding with Fock potentials provides a non-approximate inclusion of the environment effects at the HF level and can be used as a reference for other embedding techniques.

In order to apply the embedding with full Fock potentials in covalently bonded systems, the cap-free FJK approach with heterolytical bond fission was proposed. The procedure recovers the description of the total system by several monomer cycles and converges to the full HF solution in the limit of self-consistency. Solely defined on the basis of embedded monomer calculations, FJK was capable to smoothly converge towards the full HF result in terms of energy and electronic density with an increasing number of macro-iterations. The choice of charged fragments by heterolytical bond fission did not raise particular issues as the absolute HF energy was accurately reproduced within less than $3 \text{ kJ} \cdot \text{mol}^{-1}$ after four macro-iterations for most of the investigated benchmark systems. Larger errors were only found at this level for the benzene molecule, which was used to test the FJK method under extreme conditions via application to a completely delocalized system. Even though more macro-iterations are needed, the calculation of the benzene fragments still converges to the full HF treatment yielding accurate absolute and relative energies of different conformers. The delocalization character of the molecule is incrementally recovered within several monomer cycles as fragment densities are allowed to spread to the embedding regions.

Within the convergence analysis of the electronic density it was observed that the main effect of the embedding potentials is already covered after a few SCF micro-iterations within each fragment calculation. The idea to restrict the performed number of micro-iteration steps was pursued for the benchmark systems of benzophenone and alanine dipeptide. The results of the absolute and relative energy differences compared to full HF were promising as only slightly larger errors were introduced with a limited number of SCF steps per fragment, while the convergence power of the procedure was preserved.

The initial FJK method was performed using the full molecular basis set for all fragment calculations involved. This prevented the single fragments to become independent of the total system size which is a requirement

to obtain a near-linear scaling fragment approach. Therefore, a mixed basis FJK (MBFK) method was proposed which represents the embedding potential in a minimal basis set, while the large set of basis functions is only used for the calculated fragment. Additional steps of projection were needed to transfer the potentials of one fragment into the basis set of the other. A final full HF step was performed to obtain MBFJK energies and densities that are free from imbalanced descriptions of embedded fragments in different basis sets. Even though the absolute and relative energy results were comparable to the initial FJK approach, a rough estimation of the computational cost of the involved fragment calculations could show that MBFJK will not scale favorably for all molecular system sizes.

On the basis of the fragment calculations a specific area of favorable scaling compared to the full HF treatment is expected, which will become very narrow when taking into account the additional projecting operations and the final full HF step as well as an increasing number of macro-iterations. The MBFJK approach ensured a slower rise of the computational cost of each fragment calculation than the initial FJK procedure, but did not make it independent of the total system size due to the increasing number of basis functions used for potential representation. A description of the electronic densities of the embedding fragments is necessary to include their Fock potentials. Thus, no overall favorable scaling of the FJK method compared to the full HF calculation can be achieved with all remaining fragments represented by Fock potentials in each monomer calculation. A possible approximation to break this limitation is to truncate the embedding description with full Fock potentials at a certain distance from the calculated fragment and represent the long-range interacting monomers by simple point charge models. The possibility to describe distant fragments with more approximate potentials without introduction of major errors was also indicated by the application of full Fock embedding in QM/MM structures. For the FJK approach, such a truncated treatment ensures that basis functions are only needed in a restricted region around each fragment, which eventually becomes independent of the full molecular size in the case of sufficiently large systems.

Another application of full Fock potentials for embedded HF calculations was presented in setting it up as a reference for the analysis of other commonly used embedding schemes. The effect of an embedding environment on a calculated monomer density was analyzed in snapshots from different QM/MM studies. A first molecular shell around a central unit was represented by QM/MM point charge models and the resulting density and consecutive correlation energy were compared to the ones from full Fock embedding. By application of embedding with a full Coulomb potential, information on the effect of the missing exchange contributions were obtained. Especially for strongly interacting systems the point charge models were incapable of covering the entire effect of the surrounding Fock

potentials. The simple model of point charges located at the embedding atoms is insufficient to describe directed molecular interactions like hydrogen bonds which may result in differences of the correlation energy of more than $10 \text{ kJ} \cdot \text{mol}^{-1}$.

Less significant effects were found considering the impact of a secondary embedding molecular shell on the monomer density and the influence of nearby molecules on properties of higher-order. The monomer density of an ammonia molecule in the corresponding crystal structure was found to be insensitive to the embedding potential used for inclusion of a secondary monomer layer. Furthermore, this trend was also found for molecular dimer calculations in a formamide crystal structure and a butanone solute in toluene, considering their dispersion interaction energy.

For the formamide crystal structure the effect on the dispersion interaction energy changed by several orders of magnitude from the free dimer calculation to the one embedded in full Fock potentials. Nevertheless, the overall impact on the absolute value was not significant for both dimer calculations leading to changes in the dispersion interaction energy within the intrinsic error of the computational method applied. Combining these results indicates that nearby monomers have the largest impact on first-order molecular properties. The effect on higher-order properties was shown to be significantly smaller on the example of the dispersion interaction energy. However, this strongly depends on the investigated property and larger impacts are expected for quantities like the molecular dipole moment.

The analysis of the QM/MM studies revealed concepts that are well known in the context of molecular embedding. Monomers that are close to the calculated QM region have the strongest impact on the resulting first-order properties. A representation of these units via point charge models may lead to a poor description of the embedding potential due to over-polarization effects. Nevertheless, the influence of monomers further away from the QM region is significantly smaller and thus a representation by simple point charges is capable of covering the main contributions of the exact potential. For the accurate description of molecular properties inside an embedding environment, the most sophisticated model needs to be used for the monomers in the direct proximity, while more approximate potentials are sufficient to represent long-range contributions. This idea is followed by established multi-layer approaches, which describe several shells of embedding around the central QM region with schemes of lowering computational effort.

4 Local Density Fitting Approximations

The most time-consuming step in a Hartree-Fock calculation is the evaluation of the two-electron integrals of Coulomb and exchange contributions for the construction of the Fock matrix. Density fitting approximations reduce the effort of these computations by approximate descriptions of four-center integrals decomposed into two- and three-index integrals. However, due to the increased number of individual integral terms, the overall exponential scaling of the method with growing system size persists. This becomes a critical issue considering the impressive developments of post-HF correlation methods that exploit the local nature of dynamical correlation effects via the use of localized molecular orbitals.^[96-101] Indeed, the progress of these methods reached a point where the computation of the reference itself, the Hartree-Fock wave function, can become the time-demanding step of the calculation and thus act as a computational bottleneck.^[1-3;5;6] Thus, a set of viable approximations is needed to obtain an accelerated Hartree-Fock procedure competing with the reduced computational timings of local correlation methods without introducing major errors compared to the exact reference wave function.

It was already shown a decade ago that the application of local orbital spaces to the density fitted contributions of the Fock matrix may heavily reduce the corresponding computational effort.^[37] In more recent developments specific focus was set on the efficient evaluation of the exchange matrix^[102], leading to the eventual proposal of a local density fitting (LDF) approach for HF exchange.^[9] This approach allows for the restriction of exchange-related integral evaluations to entries of significant impact on the final exchange matrix. The involved restriction criteria are based on localized molecular orbitals, which are capable of exploiting the comparably short-range nature of exchange effects.

The LDF approach is not limited to the accelerated computation of a Hartree-Fock wave function. Due to the admixture of exact HF exchange in hybrid DFT functionals, also corresponding Kohn-Sham calculations benefit from a local density fitting description of exchange contributions. The approach was already established and implemented in the MOLPRO quantum chemistry package^[103] for closed-shell Hartree-Fock calculations.^[9] This set the focus on the application to open-shell calculations with the restricted and unrestricted formulations of HF and in particular of hybrid DFT procedures.

This chapter gives an overview on the application of local density fitting approximations for the computation of the exchange matrix in restricted and unrestricted SCF calculations. After the introduction and discussion of the LDF approach, further approximations for the acceleration of the general SCF procedure are presented. HF and hybrid DFT calculations with local exchange fitting on two different type of benchmark systems

are analyzed to explore the resulting limitations and speed up gains of the introduced approximations. A chapter conclusion summarizes the possibilities of the LDF approach concerning the favorable scaling compared to the non-local SCF variants and the control of the involved energy errors through adjustable local density fitting thresholds.

4.1 Method

This section describes the characteristics of the local density fitting of the exchange matrix for SCF calculations. A consistent notation to the conventional density fitting approach as provided in section 2.1.4 was used. In contrast to the standard density fitting method, the LDF approach restricts the summations involved in the computation of exchange contributions. The use of a localized molecular orbital (LMO) basis in form of intrinsic bond orbitals (IBOs)^[20] allows for the generation of different orbital domains that are limited in the number of included functions.

The first domain arises naturally from the localization of the molecular orbitals and the resulting sparsity of the LMO coefficients. Since every LMO i is restricted to a specific region of the molecular system, only a subset of all AO basis functions located on the nearby atoms contribute significantly to its construction. The coefficients $C_{\mu i}$ at each atomic center X are assumed to be negligible and thus set to 0, if

$$\sum_{\mu \in X}^{N_{\text{bas}}^X} |C_{\mu i}|^2 \leq T_{\text{LMO}} . \quad (4.1)$$

Here, T_{LMO} is the threshold for the LMO sparsity with a default value of 10^{-6} . The remaining non-zero coefficients are readjusted by fitting the approximate LMOs to the original ones. Thus, a set of AO basis functions with significant contributions to the formation of the LMO i is found and denoted as the corresponding LMO domain $[i]_{\text{LMO}}$.

The second domain results from the restriction of the number of fitting functions to the vicinity of the LMO i . A first fitting set is constructed by inclusion of all auxiliary functions with an IBO partial charge larger than 0.2. This set is subsequently extended by addition of functions on neighboring centers with a distance smaller than R_{DF} and a number of bonds between both centers less than I_{DF} . Two atoms are considered to be bonded if their distance does not exceed 1.2 times their combined atomic radii. The simultaneous application of distance and connectivity criteria ensures that basis interactions over a chain of covalent bonds are taken into account as well as the ones from non-bonded centers which are close in space. Combination of both criteria as $R_{\text{DF}} = (2I_{\text{DF}} + 1)a_0$ reduces the domain determination to only one parameter, which is by default set to

$I_{\text{DF}} = 3$. The resulting set of fitting functions is denoted the fitting domain of the LMO as $[i]_{\text{fit}}$.

Determination of the third and last domain is achieved on the basis of the three-index integral $(A|\mu\nu)$ contributions to the resulting exchange matrix. Since the values of such an integral depends on the overlap of the AO basis functions χ_μ and χ_ν , it decays exponentially with the distance between the corresponding centers. In order to estimate which integrals will not significantly contribute to the final exchange matrix, each atomic center is assigned a sphere of a specific radius. With the AO basis functions χ_μ and χ_ν being located at different atomic centers and one of them contributing to the LMO domain $[i]_{\text{LMO}}$, the integrals $(A|\mu\nu)$ are neglected if the spheres around both atomic centers do not overlap. The size of the spheres is chosen such that if all contributions of the largest normalized basis functions located at the corresponding center are neglected, the norm changes by less than the threshold T_B . The default value was chosen to be $T_B = 10^{-5}$, below which no significant improvements of the resulting accuracy were observed. All atomic functions that survived the screening with the basis functions of the LMO domain $[i]_{\text{LMO}}$ are attributed to the AO domain $[i]_{\text{AO}}$.

All three domain restrictions at first effect the half-transformation of the three-index integrals as seen for the conventional DF approach in equation (2.53). In the LDF approach this half-transformation is limited to functions included in the atomic orbital $[i]_{\text{AO}}$, local molecular orbital $[i]_{\text{LMO}}$ and fitting domains $[i]_{\text{fit}}$ of the LMO i and writes as

$$(A|\mu i) = \sum_{\nu \in [i]_{\text{LMO}}}^{N_{\text{bas}}^{\text{dom}}} (A|\mu\nu) C_{\nu i}, \quad \mu \in [i]_{\text{AO}}, A \in [i]_{\text{fit}}. \quad (4.2)$$

For every fitting domain a triangular matrix G_{AB} is determined by Cholesky decomposition of the fitting matrix Q_{AB} known from equation (2.47) as

$$[\mathbf{Q}]_{AB} = [\mathbf{G}\mathbf{G}^\dagger]_{AB}, \quad A, B \in [i]_{\text{fit}}. \quad (4.3)$$

G_{AB} is applied to the half-transformed three-index integrals resulting in a set of linear equations of the form

$$(A|\mu i) = \sum_{B \in [i]_{\text{fit}}}^{N_{\text{fit}}^{\text{dom}}} G_{AB} (\bar{B}|\mu i), \quad A, B \in [i]_{\text{fit}}, \mu \in [i]_{\text{AO}}. \quad (4.4)$$

This linear equation system is solved for each LMO i yielding the integrals $(\bar{B}|\mu i)$, which are finally used for the construction of the approximate

exchange matrix according to

$$K_{\mu\nu} \approx \sum_{i=1}^{N_{\text{LMO}}} \left[\sum_{\bar{B} \in [i]_{\text{fit}}}^{N_{\text{fit}}^{\text{dom}}} (\bar{B}|\mu i)(\bar{B}|\nu i) \right], \quad \mu, \nu \in [i]_{\text{AO}}. \quad (4.5)$$

The approximate Coulomb matrix is obtained by the conventional density fitting procedure as described in section 2.1.4. A final calculation of the energy with full fitting domains is performed once the SCF procedure converged. This significantly improves the resulting accuracy and can be carried out without the demanding computation of a new Fock or Kohn-Sham matrix.^[37]

No particular issues arise from the application of the LDF procedure in restricted open-shell or unrestricted HF calculations. While separate calls of the program are necessary for the open- and closed-shell parts of the Fock matrix in ROHF, the procedure is applied twice for the independent α and β spin treatment in the UHF formalism. The application in Kohn-Sham calculations with hybrid functionals is straightforward, since the admixed exchange originates from Hartree-Fock theory. A fully parallel version of the LDF approach with abelian point group symmetry for restricted and unrestricted HF and hybrid DFT calculations was implemented in the MOLPRO quantum chemistry package.^[103]

4.2 SCF Approximations

Two additional approximations have been introduced to accelerate critical operations of the SCF procedure and thus improve the overall speed up. These tools were applied in both, the local and conventional density fitting approach, throughout the entire course of this chapter. This provided accelerated SCF calculations, in which the speed ups and errors solely arising from the local fitting of exchange contributions could be estimated.

Starting Guess

The success of any self-consistent field approach, in which molecular orbitals are successively optimized, is depending on the quality of the initial starting guess used for a first orbital description. A straightforward starting guess can be obtained by diagonalization of the one-electron matrix \mathbf{h} and usage of the resulting eigenvectors as initial orbital coefficients. Even though this approach is direct and parameter-free, the starting orbitals yield poor electronic densities and thus a large number of SCF steps is needed until convergence is reached.

A more sophisticated starting guess is obtained by the description of superimposed atomic densities as used by default in the MOLPRO quantum chemistry package.^[103] The coefficients of the used AO basis functions

are chosen to reproduce atomic densities with occupation numbers of the valence orbitals stored in a library for each atom type. This approach corresponds to the description of neutral and spherical atoms within the system and yields starting densities that are in most cases already quite close to the actual molecular one.

However, the computation of this guess can be rather demanding in the case of large basis sets. By the application of a minimal set of AO functions at each atom, the coefficients of the starting orbitals are quickly determined. If cc-pVnZ Dunning basis sets^[67] are used for the SCF calculation, the corresponding minimal basis is simply constructed from a subset spanning the minimum of AO space for each atom. For other basis sets, such as the Pople or Karlsruhe type^[104;105], the minimal orbital basis set stored in a library for a comparable cc-pVnZ size is applied to create the initial guess. Their transfer to the actual basis set is finally achieved by orbital projection.

This approach ensures a fast computation of the starting orbital guess, even for large systems with extended basis set treatment. As long as the initial density description is not too far from the actual molecular situation, a fast and stable convergence of the SCF procedure is expected for the majority of systems. Since the starting density matrix will be extremely sparse, a quick computation of the first Fock or Kohn-Sham matrix is achieved. Especially in the case of a large number of basis functions used for the SCF calculation, this will lead to significant computational savings in the evaluation of the starting guess.

Matrix Diagonalization

One of the demanding operations during each SCF step is the diagonalization of the entire Fock or Kohn-Sham matrix in HF and DFT calculations, respectively. In the local and conventional density fitting approaches of this work, the full matrix diagonalization was only carried out in the first few iterations and in the final non-local step of the SCF cycle. If during the procedure the total energy was not changing more than $0.01 E_h$ compared to the previous SCF step, an efficient pseudodiagonalization procedure was applied to reduce the computational cost of this algebraic operation.^[37]

This approach diagonalizes the matrix of interest exactly only in an orbital space of a $\pm 1 E_h$ range below the highest occupied molecular orbital (HOMO) and above the lowest unoccupied molecular orbital (LUMO). An approximate technique is performed to achieve near-diagonalization of the remaining orbital space by simple orbital rotations. Thus, the region of interest for chemical application, the valence orbitals and frontier orbitals of the virtual space, are treated exactly, while the matrix diagonalization operation is accelerated by approximations effecting the core and high-lying virtual orbitals. This approach drastically reduces the employed

CPU time for matrix diagonalization per SCF step without introduction of significant errors to the final results or to the convergence of the complete SCF procedure.

4.3 Benchmark Calculations

The local density fitting approach was applied in the restricted and unrestricted SCF formulations to calculate open-shell systems. Two benchmark sets were designed to test LDF in HF and hybrid DFT calculations with different purposes in mind. On the one hand the limitations of the method are explored in highly delocalized systems, on the other hand one is interested in the critical molecular size from which on the local fitting of exchange contributions lead to a significant speed up compared to the established non-local density fitting procedure.

A first set of benchmark systems contained systems of pronounced π -conjugation character. These structures allow for the analysis of the LDF performance in extreme cases and provide insights into the control offered by the threshold parameters in the total or relative energy errors.

In a second set of test calculations, RAFT initiation reactions were considered including the formation of different radical intermediates. This reflects a more realistic application to analyze the errors in relative energies and the critical molecular size after which the local approximations in fact lead to computational savings.

4.3.1 Delocalized Systems

Structures containing an unpaired electron which is strongly delocalized among the system, represent a natural challenge for the application of local orbital approximations. With an increasing π -character and conjugation of such systems the HOMO-LUMO gap rapidly decreases moving the occupied and virtual orbital spaces closer together. At a certain point of delocalization, the involved orbital interactions will become so significant that a reliable orbital localization is no longer possible. In this limit the description of the systems with local density fitting approximation will eventually break down.

In order to explore such delocalization limits, two types of structures with an adjustable level of π -conjugation were analyzed. The first set of molecules were inspired by the radical intermediate during the conversion of polyunsaturated fatty acids. A hydrogen atom was removed from the central position of an alkane chain containing 17 carbon atoms and the number of double bonds within the system was varied to achieve different degrees of delocalization. Four symmetrical structures **1A-1D** were generated containing 2, 4, 6 and 8 double bonds in the heptadecanyl radical chain, as shown on the left side of Figure 4.1.

The second set of molecules is represented by the classical players concerning delocalized systems in form of cycloaromatic compounds. Since the analysis of open-shell applications was desired, an electron was removed from each structure leading to cationic systems. The increasing delocalization character was achieved by a growing system size among the cations of benzene, naphthalene, anthracene and tetracene. These structures **2A-2D** are displayed in the right part of Figure 4.1.

Both molecular sets were optimized with C_{2v} point group symmetry at the ROB3LYP/cc-pVDZ level of theory. Afterwards, the LDF approach in HF and Kohn-Sham (KS) hybrid DFT calculations with the B3LYP functional^[24;25] was applied. The restricted open-shell and unrestricted formalism was used for both SCF procedures with a cc-pVTZ^[67] basis set and the corresponding set of auxiliary functions.^[68] Default criteria established for the closed-shell LDF applications of LMO sparsity $T_{LMO} = 10^{-6}$ and fitting domain size $I_{DF} = 3$ were applied. All energies were compared to the ones from their non-local density fitted variant and the resulting error was divided by the number of carbon atoms n_C to remove the size-dependency. Thus, the errors solely arising from the local exchange fitting approximation could be analyzed free from effects due to an increasing system size in the case of the structures **2A-2D**. The corresponding results for both test sets are displayed in the top two panels of Figure 4.2.

Looking at the results of the systems **2A-2D** and **1A** it seems that the effect of the LDF approximations on the resulting error is less pronounced for the KS formalism than for HF. All five systems exhibit smaller errors compared to their non-local treatment if restricted open-shell and unrestricted DFT calculations are performed, while larger errors are found in the corresponding HF applications. However, the same is no longer true for **1B-1D**, where the resulting errors per carbon atom are rather similar for

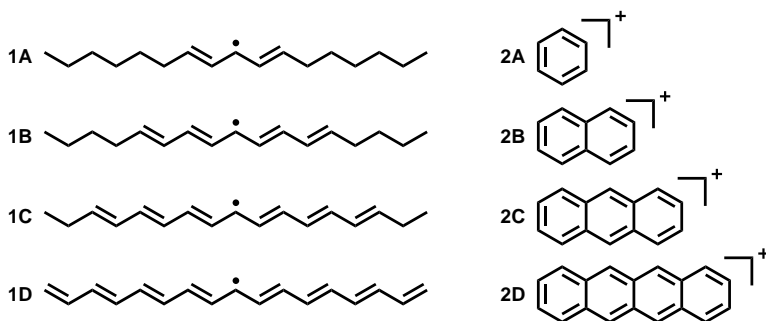


Figure 4.1: Benchmark systems included in the analysis of open-shell local density fitted HF and hybrid DFT calculations. **1A-1D**: Heptadecanyl radical with an increasing number of double bonds and delocalization character along the system. **2A-2D**: Cations of benzene, naphthalene, anthracene and tetracene.

the used methods. Thus, the results are not significant to deduct a reliable trend on the effect of local density fitting among the different procedures applied.

For both sets, **1** and **2**, the rising delocalization character going from molecule **A** to **D** leads to an increase of the corresponding energy error per carbon atom. This is perfectly expected due to the hampered orbital localization in systems of strong π -conjugation. Most chemical applications will feature delocalized systems comparable to **1A-1B** and **2A-2B**, in which the errors per carbon atom are still below $10^{-7} E_h$. This strengthens the initial hypothesis that the same default LDF criteria as in the closed-shell

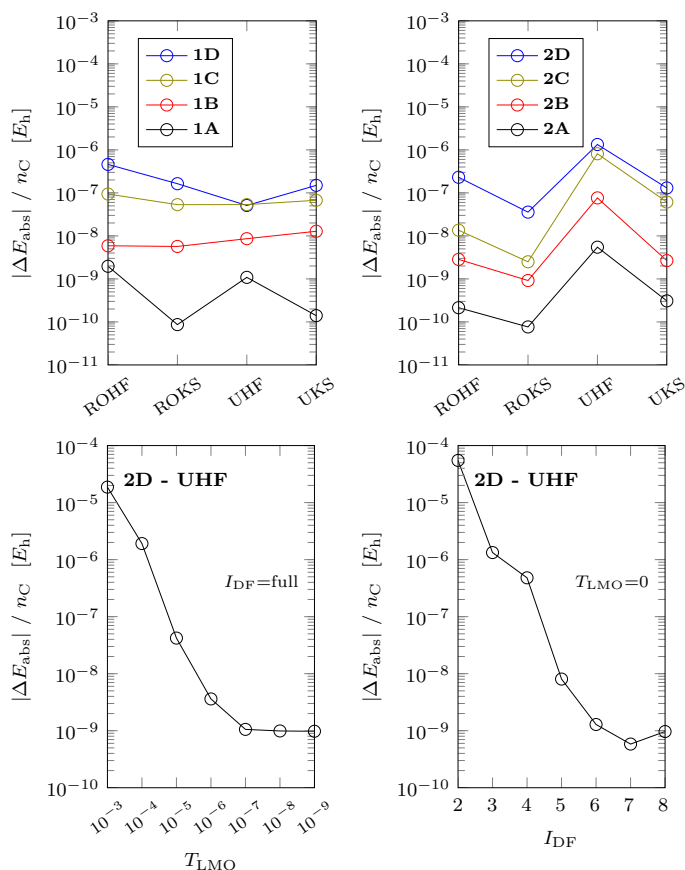


Figure 4.2: Total energy errors per number of carbons of the local density fitting approach compared to the non-local variant. Restricted and unrestricted open-shell HF and KS (B3LYP) results for the systems **1A-1D** (top left) and for the systems **2A-2D** (top right). UHF results for the system **2D** with varying LMO sparsity T_{LMO} (bottom left) and varying fitting domain size I_{DF} (bottom right).

case can be applied for open-shell calculations. Even in case of the largest error, the UHF treatment of tetracene **2D**, the difference to the non-local procedure is only slightly above $10^{-6} E_h$ per carbon atom. Although this impact of the local density fitting approximation on the results is not negligible anymore, it is still below the total effect of the conventional density fitting approach on the Hartree-Fock energy. When relative energies are considered, as commonly done in chemical applications, the error resulting from the LDF approach will be partly removed. Furthermore, the errors in the absolute energies can be reduced by the application of stricter criteria within the local density fitting procedure. This is analyzed in detail for the UHF calculation of tetracene **2D**.

The UHF treatment of system **2D** was repeated with different criteria of LMO sparsity T_{LMO} and fitting domain size I_{DF} . While one parameter was varied to obtain the energy difference to the non-local DF procedure, the other one was set to its optimum value including either the full fitting domains for different LMO sparsities or the full LMO domains with altering fitting domain size. The results are shown in the bottom two panels of Figure 4.2.

Starting from loose LDF criteria large errors are observed in the region of $10^{-5} E_h$ to $10^{-4} E_h$ due to the limited size of either the LMO or the fitting domains. Varying the parameters of the local exchange fitting to a lower threshold in LMO sparsity or a larger number of bonds for fitting domain formation leads to smooth convergence with an error limit around $10^{-9} E_h$ per carbon atom. Thus, the errors introduced by the LDF approach can be erased to an adjustable level by modification of the criteria applied.

Both criteria analyses can be used to estimate the effect of the complete inclusion of the LMO or fitting domains, while the parameter of the other domain remains at its default value. Adding the full LMO domains in an UHF calculation with the default fitting domain parameter of $I_{DF} = 3$ leads to an error in the region of $10^{-6} E_h$, while the full fitting domain inclusion for the default parameter of $T_{LMO} = 10^{-6}$ yields a difference between $10^{-9} E_h$ and $10^{-8} E_h$ per carbon atom. This indicates that the error introduced by local fitting of exchange contributions is in general more sensitive to the size of the fitting domains than to the LMO sparsity.

4.3.2 RAFT Polymerization

The reversible addition-fragmentation chain transfer (RAFT) polymerization process yields macromolecular polymers of controllable chain length.^[106] In an precursory step of the polymerization procedure a chosen RAFT agent is activated with a radical starter molecule forming a pre-equilibrium with the resulting radical product. Due to the large variety of available RAFT agents and radical starters for polymerization, this pre-equilibrium reaction is perfectly suited for the analysis of local density fitting approximations

in open-shell systems of variable size.

The starting reactions of five different RAFT agents following several works on RAFT polymerization processes were chosen for LDF analysis:

- 2-(2'-cyanopropyl) dithiobenzoate (CPDB)^[107]
- 1-phenylethyl dithiobenzoate (PEDB)^[107]
- 4-cyanopentanoic acid dithiobenzoate (CPADB)^[108]
- *S,S'*-bis(methyl-2-propionate) trithiocarbonate (BMPT)^[109]
- *S*-dodecyl-*S'*-(α,α' -dimethyl- α'' -acetic acid) trithiocarbonate (DDMAT)^[110]

Pairing the above-listed RAFT agents with the radical starters used in the

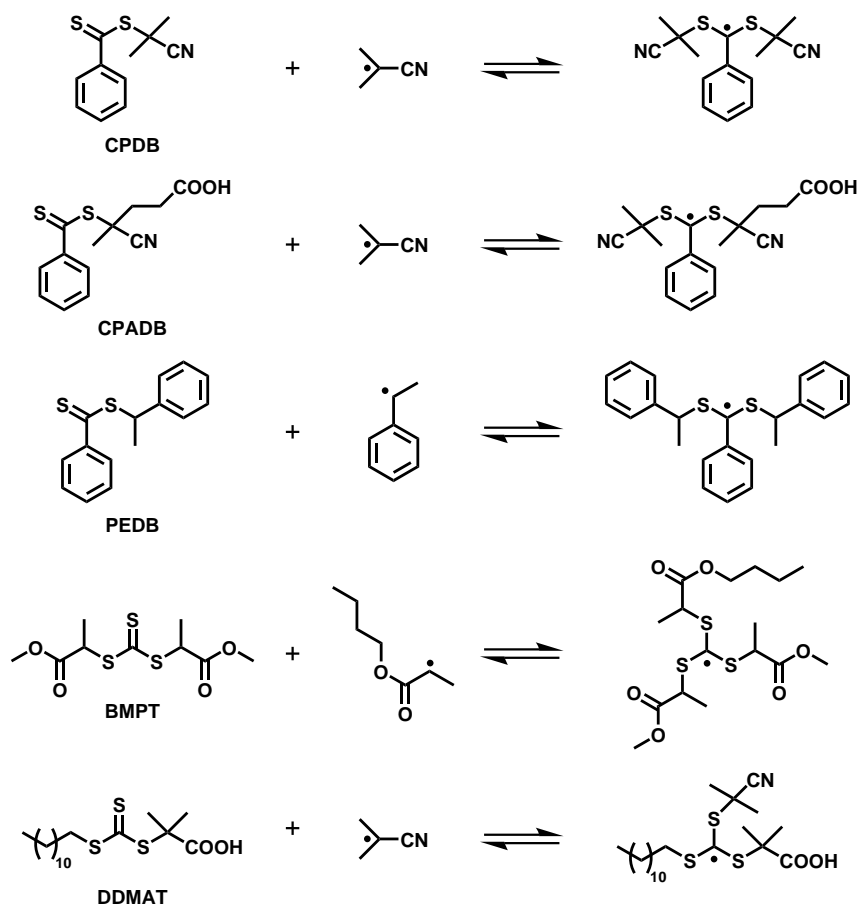


Figure 4.3: Benchmark set of RAFT polymerization initiation reactions used for comparison of the local and non-local density fitted open-shell HF and DFT methods. The reactions are sorted according to an increasing size of the product radical.

corresponding studies yielded an increasing size of the resulting radical product reaching from 36 to 65 atoms. According to this criteria the reactions of all five RAFT pre-equilibria are listed in Figure 4.3.

The structures were optimized at the B3LYP/def2-TZVP level of theory with the restricted formalism for the closed-shell RAFT agents and the unrestricted formulation for the radical structures. Solvation effects of toluene were taken into account during the optimization with the COSMO solvation model using $\epsilon = 1.89$.^[111] Local density fitting approximations were applied to the calculation of all compounds involved in the analyzed RAFT reactions. Restricted open-shell HF and unrestricted Kohn-Sham DFT with the B3LYP functional^[24;25] were used for the calculation of the radical species of the starter molecules and the resulting products. The non-radical RAFT agents were treated with the restricted closed-shell formulations of HF and DFT. All calculations were performed several times using Dunning basis sets of double- ζ , triple- ζ and quadruple- ζ quality in cc-pVDZ, cc-pVTZ and cc-pVQZ with density fitting basis sets of corresponding size.^[67;68] The results were compared to the non-local density fitted variants of treatment in order to quantify the occurring errors in relative energies and the speed up gained by the local fitting of exchange contributions for different system sizes. Table 4.1 contains information on the number of AO and fitting basis functions, the average fitting domain size and the speed ups compared to non-local density fitting for all products of the RAFT reactions considered.

Since the experimentally accessible equilibrium constant of the RAFT reactions depends exponentially on the relative energy between both reaction sides, the accurate description of the electronic reaction energy $\Delta E_{\text{reac}} = E_{\text{prod}} - (E_{\text{agent}} + E_{\text{starter}})$ is of critical importance. In the considered RAFT reactions the errors in the electronic reaction energies compared to the non-local calculation of all compounds were found to be extremely low for the HF and DFT treatment with three different basis sets. Using the default local density fitting criteria of $T_{\text{LMO}} = 10^{-6}$ and $I_{\text{DF}} = 3$, even the largest error of ΔE_{reac} did not exceed $0.002 \text{ kJ}\cdot\text{mol}^{-1}$, which is below the intrinsic error of density fitting approximations.^[112]

The data gathered in Table 4.1 focuses on the applied basis set sizes and the computational savings obtained per iteration and compared to the full SCF calculation. All values refer to the radical products of the RAFT reactions yielding a range of 359 up to 2637 AO basis functions, with a maximum fitting basis of 4926 functions. The parameter $q_{\text{fit}} = \langle N_{\text{fit}}^{\text{dom}} \rangle / N_{\text{fit}}$ represents the average number of functions per fitting domain divided by the total size of the auxiliary density fitting set. Larger computational savings are expected for lower values of q_{fit} , since on average a smaller amount of fitting functions of the entirety are used to form the density fitting domains.

Considering the intermediate basis set size of cc-pVTZ, one can observe

Table 4.1: Speed ups of the density fitted restricted open-shell HF (ROHF) and unrestricted Kohn-Sham (UKS) approach with the B3LYP functional compared to their non-local variants for a set of RAFT polymerization initiation reactions. The number of AO basis functions N_{bas} , fitting functions N_{fit} and the average fitting domain size divided by the total number of fitting functions q_{fit} are provided. f is the speed up factor of the local and non-local procedures given for the calculation of the reaction product in total (f_{tot}) and in average per SCF iteration ($\langle f_{\text{iter}} \rangle$) calculated using 4 cores (Intel 3.1Ghz).

System	N_{bas}	N_{fit}	q_{fit}	(L)DF-ROHF		(L)DF-UKS	
				$\langle f_{\text{iter}} \rangle$	f_{tot}	$\langle f_{\text{iter}} \rangle$	f_{tot}
<u>cc-pVDZ</u>							
CPDB	359	1805	0.51	0.83	0.84	0.89	0.90
CPADB	425	2131	0.43	0.90	0.92	1.08	1.03
PEDB	473	2363	0.38	1.06	0.99	1.26	1.20
BMPT	497	2497	0.34	1.19	1.16	1.41	1.31
DDMAT	580	2890	0.30	1.60	1.50	1.92	1.72
<u>cc-pVTZ</u>							
CPDB	816	2095	0.51	0.96	0.95	1.09	1.06
CPADB	964	2471	0.43	1.11	1.10	1.33	1.28
PEDB	1080	2749	0.37	1.30	1.28	1.45	1.36
BMPT	1140	2911	0.33	1.49	1.44	1.74	1.62
DDMAT	1354	3399	0.30	2.02	1.90	2.31	2.09
<u>cc-pVQZ</u>							
CPDB	1563	2965	0.50	1.13	1.13	1.26	1.22
CPADB	1843	3491	0.42	1.31	1.29	1.44	1.38
PEDB	2073	3907	0.37	1.46	1.41	1.61	1.52
BMPT	2197	4153	0.32	1.68	1.59	1.90	1.76
DDMAT	2637	4926	0.29	2.44	2.29	2.80	2.52

that for the smallest radical RAFT product, the one formed from the CPDB agent, the total LDF calculation can even be slightly slower than the conventional DF run which is indicated by a speed up factor of $f_{\text{tot}} < 1$. The corresponding system contains only 36 atoms and is thus barely touching the turning point of a beneficial LDF scaling. For the smaller cc-pVDZ basis the observation of an inferior scaling compared to the non-local DF is made even for slightly larger RAFT products. The orbital

localization and the computation of the individual local density fitted exchange contributions dominate the computational timing in systems of such limited basis set size. In the case of the increased cc-pVQZ basis set, the turning point is reached for all systems yielding speed ups that are favorable compared to conventional DF procedures in terms of the total calculation and also on average per SCF iteration.

The proportion between the average fitting domains and the total fitting basis set size indicated by q_{fit} is quite constant for each system among the different basis sets used. This indicates that q_{fit} is mostly dependent on the electronic structure and the molecular geometry, not on the number of basis functions per site. Consistent to this dependency is the decreasing value of q_{fit} with an increasing system size from the radical product of CPDB to the one from DDMAT. The speed up gained by the LDF approach goes hand in hand with a smaller q_{fit} and a rising number of basis functions. This is seen at its maximum in the UKS calculation of the DDMAT radical product with a cc-pVQZ basis set yielding a speed up factor of 2.52 in the total timing and even 2.80 for the individual SCF iterations.

In all but the smallest analyzed structures, the average speed up per SCF iteration is larger than for the corresponding total HF or DFT calculation. This difference arises from the initial orbital localization and the final non-local SCF step performed in the LDF procedure. A stronger discrepancy of both speed up factors is found for the Kohn-Sham DFT calculations than for the Hartree-Fock treatment. In the analyzed radical products also the general speed up gain of the DFT calculations is more significant than the ones of the performed HF for all basis set and system sizes considered. Since the evaluation of exchange contributions is the computationally demanding step in both, Hartree-Fock and hybrid DFT approaches, the speed up differences are mainly attributed to the applied formalisms. An unrestricted Kohn-Sham hybrid DFT approach was carried out in which the exchange contributions of α and β electrons are calculated independently. While the applied ROHF procedure performs only the localization of closed-shell and open-shell orbitals separately, their exchange contributions are evaluated by a single call of the local density fitting routine. In contrast to this the unrestricted formalism requires two calls of the LDF procedures and thus a larger speed up compared to the conventional density fitting approach is observed.

The analyzed RAFT benchmark can be used to identify the onset of computational savings achieved by the LDF algorithm. With a standard triple- ζ basis, the turning point of a beneficial scaling in ROHF was reached between the radical products of CPDB and CPADB, two systems with 36 and 42 atoms. For the larger cc-pVQZ basis set the favorable scaling was already achieved for the smallest system, while it was shifted between the products of PEDB and BMPT with 48 and 52 atoms for the cc-pVDZ basis set. However, beyond this turning point a constant increase of the

speed up gained by LDF approximations is observed making the approach applicable to rapid calculations of large systems with errors controllable by the thresholds used for domain formation.

4.4 Conclusion

The local density fitting approximation was applied to the calculation of different benchmark systems in restricted and unrestricted HF and hybrid DFT formulations. Several properties of the approach were investigated including its limitations in strongly delocalized systems, the crossing point towards a favorable scaling compared to conventional density fitting and its accuracy in terms of absolute and relative electronic energies. The possible application of the approximation for larger systems was estimated on the basis of smaller benchmark calculations by analysis of the controllability of the resulting errors and the gain in computational speed up.

In the case of systems with a pronounced π -conjugation character, the local fitting of exchange contributions was pushed to its limits. The resulting energy error increased with a rising degree of electron delocalization due to the poor description of the conjugation effect via local orbital spaces. However, even for the system with the highest error in the electronic energy, the description could be smoothly improved by adjustment of the thresholds of the LMO sparsity T_{LMO} and the fitting domain size I_{DF} . By increasing the average number of fitting functions per domain or the allowance of larger LMO domains, the description of the system can be successively improved towards the non-local limit with errors in the region of $10^{-9} E_{\text{h}}$. Thus, the errors introduced by the local density fitting approximation can be strictly controlled in terms of the involved thresholds for the different domains, whereas the size of the fitting domains were found to have a larger impact on the resulting energy than the LMO sparsity.

The critical molecular size from upon which the LDF approximations lead to a beneficial speed up compared to the conventional density fitting was estimated for a set of RAFT pre-equilibrium reactions. By application to the radical products of the RAFT reactions, a suitable window for the local density fitting analysis was found. The involved structures covered the cases of an unfavorable scaling compared to the standard DF approach, moderate speed ups and significant savings of computational time by a total factor of 2 or more. The critical molecular size of a beneficial scaling could be estimated for system sizes between 36 and 42 atoms for a standard cc-pVTZ basis set. In the case of the smaller cc-pVDZ basis the turning point for this behavior was shifted between 48 and 50 atoms. If a large basis set like cc-pVQZ was applied, the crossing in computational scaling was already reached for all of the investigated systems and thus had to be located in a region below 36 atoms.

The investigated RAFT reactions also showed an increase of the speed

up gain of local exchange fitting with an enlargement of the used basis set. Going from cc-pVDZ over cc-pVTZ to cc-pVQZ, the speed up factors of the total calculation and per SCF iteration significantly increased with a larger number of AO basis and fitting functions used for the computation of the same structure. The average number of fitting functions divided by the total fitting basis size q_{fit} was found to be constant among the applied basis sets. This indicated that the changes of q_{fit} arise according to the molecular size and not according to the number of basis functions per atom. Within an applied basis set the speed up gain was obviously rising due to the increasing system size and the larger number of AO basis and auxiliary fitting functions involved.

Speed up gains provided by the LDF approach were analyzed in terms of the acceleration of the total SCF procedure and on average per iteration within the corresponding cycle. The general trend exhibited a larger computational saving per SCF iteration than for the entire calculation, since additional operations such as the orbital localization and the final non-local SCF step are required within the LDF approach. An estimate of the impact of the local exchange fitting on HF and hybrid DFT calculations could also be performed on the basis of the RAFT benchmark results. While generally improved speed ups were obtained for the performed Kohn-Sham hybrid DFT calculations, the direct comparison to the Hartree-Fock results was impossible due to the different formalisms in form of restricted and unrestricted formulations used. A similar effect for both SCF procedures is expected due to the inclusion of the same type of exchange contributions, but the influence was superimposed by the advantageous speed up gain from a double call of the LDF routine in the unrestricted open-shell treatment.

The overall acceleration of the SCF procedure by local exchange fitting with controllable errors encourages the application on medium-sized to large systems. With significant speed up gains for a moderate basis set size already resulting in systems of roughly 50 atoms, the LDF approach allows for the performance of otherwise demanding computations in open-shell systems. Hybrid DFT calculations are the method of choice for the calculation of specific areas in biomolecular systems but are still limited to medium-sized structures due to the steep scaling of the involved HF exchange. The local density fitting of exchange is capable of breaking this limitation and may provide a more complete picture of larger biomolecular regions in a reasonable computational time. This idea will be pursued through application of the LDF approach on a substrate docking to the active site of an enzyme in the following chapter, dealing with a large open-shell system far beyond the 50-atom range.

5 Theoretical Study of CspLOX2 Selectivity

Enzyme studies are perfectly suited for the application of local and fragment potential calculations. The enzyme catalysis takes place in a specific region, the active site, which is of sufficiently limited size to employ quantum mechanical methods. This allows for the simulation of the reaction via small model systems. However, due to its specific interactions and the long-range Coulomb effect involved^[113;114], the surrounding protein environment can have a huge impact on the catalyzed chemical transformation. Due to the sheer enzyme size, the main part of the system can only be treated at the molecular mechanics level as it is commonly done in QM/MM approaches.^[74;78;79] In order to enlarge the investigated quantum mechanical region, effective potentials of the nearby environment or local density fitting approximations are useful approaches, providing the possibility to study enzyme catalysis in the active site under the effect of the embedding protein.

This chapter starts with an introduction to the class of lipoxygenase enzymes. The section on computational details of the electronic structure and molecular mechanics calculations is followed by the different factors regarding the enzymatic selectivity of the CspLOX2 system. A chapter conclusion with some final remarks on the methods and obtained results is given at the end.

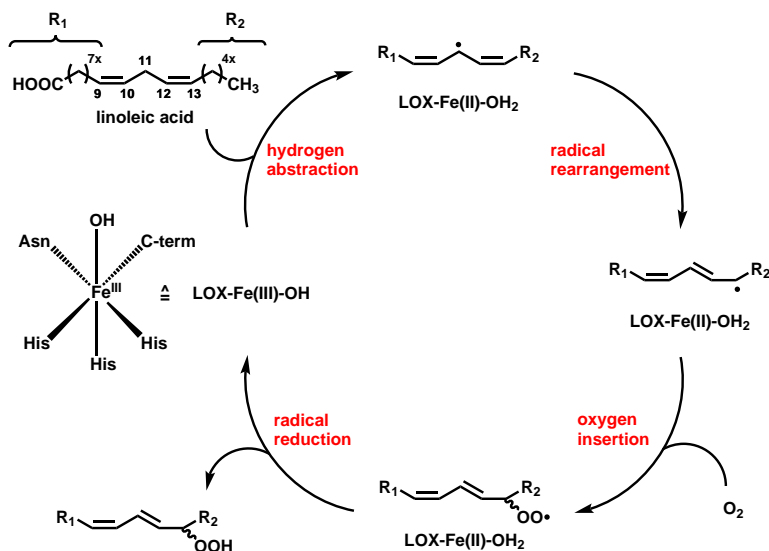


Figure 5.1: Catalytic cycle of the non-heme iron lipoxygenase (LOX) class of enzymes for the linoleic acid (C18:2) substrate. Radical rearrangement and dioxygen insertion are the steps considered in detail in this work.

5.1 Lipoxygenases

The lipoxygenase (LOX) class of enzymes plays a key role in metabolic processes for the dioxygenation of polyunsaturated fatty acids in lipids.^[10;11] Reactions catalyzed by LOXs take place ubiquitously in higher plants and animals but have also been identified in several corals, mosses, fungi and bacteria.^[115;116] The active site of these enzymes contains a non-heme iron or manganese moiety reacting with the *cis,cis*-1,4-pentadiene unit of the fatty acid substrate. In the iron case the Fe(III) metal is typically coordinated in an octahedral geometry by three histidine nitrogen donor atoms, the amide oxygen of an asparagine residue, the enzyme C-term carboxylate and a free hydroxide ligand.

In a first step of the catalytic cycle, as shown in Figure 5.1, a hydrogen atom is abstracted from the substrate and transferred to the iron bound hydroxy group yielding a water ligand and Fe(II) in high spin configuration. The substrate is converted to a bisallylic radical species that can undergo radical rearrangements as shown in the second step. These rearrangements allow for the formation of conjugated double bonds with the unpaired electron located at the outer positions of the pentadiene unit, which corresponds to the thermodynamically favored radical configuration.^[117] The third step of the cycle adds dioxygen to form a superoxide radical. Upon radical reduction in the last step a substrate hydroperoxide is formed and the LOX catalyst is restored. While the hydrogen transfer has been analyzed and well understood^[118-125], little is known about the mechanism of molecular oxygen addition.

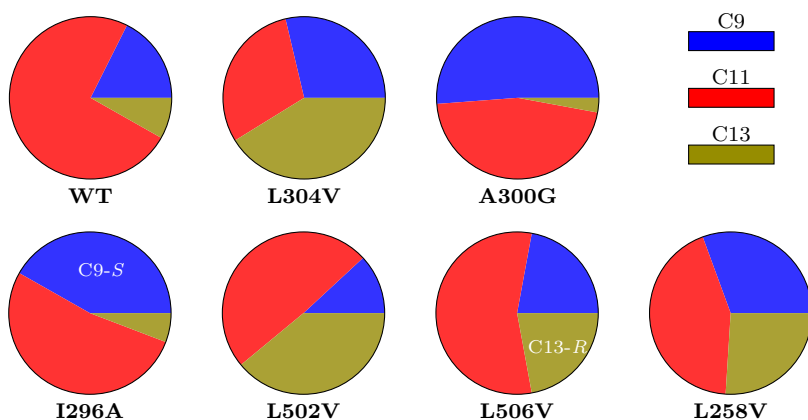


Figure 5.2: Product distributions of the CspLOX2 wild type (WT) and a set of single amino acid mutants. For mutations the first letter gives the amino acid which is replaced, whereas the number indicates the chain position and the last letter corresponds to the new amino acid. The regular main stereo products are C9-*R*, C11-*R* and C13-*S*. Changes in the preferred stereoselectivity are written inside the diagram.

By using linoleic acid (C18:2) as a substrate the LOXs enzymes catalyze the dioxygenation in a way that normally yields conjugated products with hydroperoxides located at the C9 and C13 positions of the fatty acids. However, bisallylic hydroperoxides at the C11 carbon atom were found to be the main products in manganese LOXs^[126;127] and also in the iron-containing CspLOX2 from the cyanobacterium *Cyanothece* PCC8801.^[128] The underlying mechanism of selectivity control for dioxygen addition has been the ground for broad discussions over the past 20 years.^[129–131] In the case of the CspLOX2 enzyme it was shown by the Feussner group that single amino acid mutations in the vicinity of the active site drastically change the resulting regio- and in some cases even the stereospecificity by strongly varying product distributions.^[132] The corresponding experimental product distributions are shown in Figure 5.2, while a cartoon representation of the CspLOX2 wild type structure is displayed in Figure 5.3.

The product distributions reach from clear preference of the C11-*R* as in the CspLOX2 wild type over almost equally distributed products as in the case of the L258V mutant up to C9-*R* and C13-*S* being the main products as seen for A300G and L304V. Two cases of changes in stereochemistry are observed inverting the product at C9 for the I296A mutation and at C13 regarding the L506V mutant. In order to understand the underlying enzyme selectivity electronic structure and molecular mechanics computations were carried out to test several proposed factors steering the observed specificity. For other LOXs these factors included spin localization^[133], activation

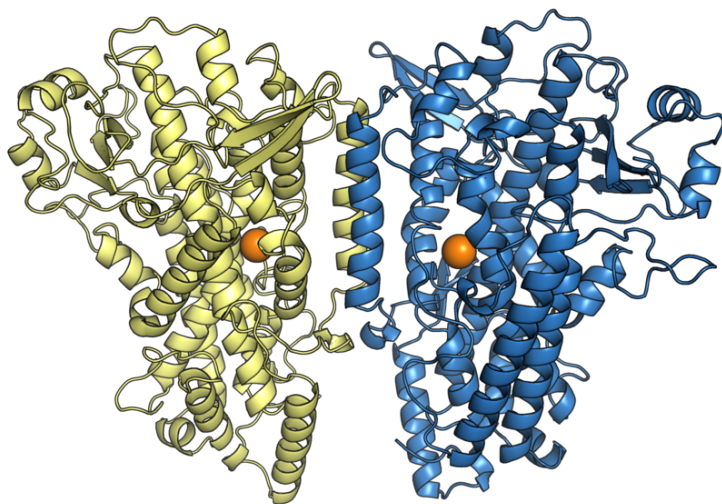


Figure 5.3: Cartoon representation of the CspLOX2 wild type enzyme. The dimeric structure is indicated by the colored amino acid chains, while the iron centers of the active sites are shown as spheres.

barriers for dioxygen addition^[134], substrate conformation, as well as steric constraints in the active site.^[135;136]

Spin localization within the substrate follows the assumption of molecular oxygen attacking the carbon site of highest spin density and provides a possible explanation for the enzymatic selectivity. The idea consists of substrate torsions induced by the enzyme which will localize the spin density of the pentadiene unit to specific carbon atoms and thus making them more reactive for dioxygen addition. A similar concept exists in the change of the activation barriers for molecular oxygen addition. Due to different heights in the activation barriers for reactions at C9, C11 and C13, the pathways are kinetically discriminated by the enzyme. Since these activation barriers may also change according to torsions of the pentadiene unit, one is in general interested in the preferred substrate conformation within the enzyme pocket. The configuration of substrate binding as well as the access to of dioxygen to the reactive carbon centers are referred to as another proposed selectivity factor in form of sterical shielding. In this concept the enzyme moves amino acid residues in front of the reactive carbons in order to mechanically block these sites for the attack of molecular oxygen attack. This would force the reaction to happen at the remaining unshielded carbons.

5.2 Computational Details

Parts of this and the following section are featured in

"Lipoxygenase 2 from *Cyanotheca* sp. controls dioxygen insertion by steric shielding and substrate fixation",

J. Newie, P. Neumann, M. Werner, R. A. Mata, R. Ficner and I. Feussner, *Sci. Rep.*, 7, 2069, 2017.

For a small model system, the 2,5-heptadienyl radical, standard unrestricted density functional theory calculations with the B3LYP functional^[24;25] and the def2-TZVP^[105] basis set (UB3LYP/def2-TZVP) were carried out with the ORCA electronic structure package.^[137] Location of the lone electron was determined by Pipek- Mezey localization of the occupied orbitals^[19], followed by a natural population analysis (NPA) of the resulting localized SOMO.^[92] The spin density distribution was scanned along an internal dihedral angle which was varied from 150° to 210° in steps of 5° with 180° corresponding to the preferred planar substrate conformation.

Activation barriers for dioxygen addition were calculated for the same model system by using the same electronic structure theory with additional damped dispersion corrections^[30;31] (UB3LYP-D3/def2-TZVP). Molecular oxygen was moved towards the carbon centers of interest in relaxed surface scans with fixed internal carbon dihedrals. The scans were performed for the same range of torsional angles as mentioned above.

The provided crystal structure of the CspLOX2 wild type dimer^[132] was used as a starting point for molecular dynamics simulations. All calculations were carried out only considering one chain of the dimer without observing any significant loss of conformation in the course of all dynamics. After cleaning up the structure, the monomer was placed in a water box with 8 Å distance to the periodic cell limits. The charge in the cell was neutralized by addition of 11 Na⁺ ions. 2000 steps of minimization were carried out with a restraint of 5 kcal/Å² placed on all non-hydrogen atoms in the enzyme, followed by 3000 steps with no restraints. All atoms were modeled with the AMBER ff10 force field.^[62] A non-bonded sphere model was used for the Fe²⁺^[138] ion and the parameters for the substrate were taken from Furse *et al.*^[136] using Merz-Kollman charges^[139] at the UB3LYP/6-31G* level of theory. Single amino acid mutations were generated by replacing the corresponding residue manually in order to simulate the L304V, A300G, I296A, L502V, L506V, L258V mutated enzymes. For the wild type and each mutation, three starting structures were prepared by placing the radical of linoleic acid in the active site pocket in different conformations. The SHAKE algorithm was used for all MD runs, with a 2 fs time step, and a 12 Å cutoff for non-bonded interactions. A Langevin thermostat was used throughout.^[64] After 50 ps of heating to 300 K, 350 ps of NPT dynamics were carried out at ambient conditions for equilibration. The production runs consisted of 1 ns for each starting structure. 5000 snapshots were used for all following analysis.

The snapshots from all simulations were taken to analyze the dihedral angle distribution of the substrate. The C10-C11-C12-C13 dihedral was extracted for all 5000 structures, resulting in a histogram of conformers in a 60° range. In order to readdress the spin density calculations within a larger enzyme environment the observed structures for the wild type were used to create three different sizes of model systems around the linoleic acid radical. Their spin density distribution was calculated with a development version of the MOLPRO2015.1 program^[103] applying local density fitting approximations as described in section 4.1 with the default parameters $I_{DF} = 3$ and $T_{LMO} = 10^{-6}$ in unrestricted density functional theory with the B3LYP functional^[24;25] and a def2-SVP^[140] basis set (LDF-UB3LYP/def2-SVP). The results were compared to the ones of the non-local density fitting procedure in terms of accuracy and computational timings.

For the large model system of the active site, containing the substrate radical, the iron moiety and a set of nearby amino acids structures, different conformers were generated for the CspLOX2 wild type and the L502V mutant from their MD simulations. The conformers were assigned to bins of 5° of the aforementioned dihedral and five random geometries were calculated in each bin. The location of the unpaired substrate electron was determined by natural population analysis of the localized SOMO as

already described above. This procedure ensured a clean separation of the substrate spin density distribution from external spin influences of the iron moiety. By averaging the results for each bin and weighting them with their conformational relevance from the observed dihedral angle distribution, the product ratio of the hydroperoxides was obtained solely on the basis of the spin density.

5.3 Selectivity Analysis

This section is guided by the proposed factors for the selectivity of the CspLOX2 enzyme, namely the analysis of the substrate spin density, activation barriers for dioxygen addition, substrate conformations and the investigation of steric shielding effects.

5.3.1 Spin Density

Changes in the spin density distribution of the substrate would provide a simple explanation for the observed selectivities with respect to the assumed molecular oxygen attack the carbon center of highest spin density.^[130;133;136] The enzyme might induce slight out-of-plane bending within the substrate, leading to a spin localization at specific carbon atoms and thus making these sites more reactive for dioxygen addition. In order to model the bound linoleic acid radical, a 2,5-heptadienyl radical model system was employed and its spin density distribution was analyzed performing UB3LYP/def2-TZVP calculations. By varying one of the dihedrals, the conjugated radical character over the whole molecule was broken and therefore the spin density was localized on the reactive carbon atoms. The amount of spin density in the SOMO at the three carbons of interest was analyzed in a dihedral range from 150° to 210°. A picture of the model system as well as the results of the spin density analysis are provided in Figure 5.4.

Starting from the preferred planar conformation corresponding to a dihedral angle of 180°, the spin density is highest at the central C11 carbon and equal amounts are located at the outer C9 and C13 sites of the pentadiene unit due to the symmetry of the model system. With variation of the dihedral angle φ , the moiety including C13 is bent out of the molecule plane. This explains the lower amounts of spin density found at this carbon site for the rest of the dihedral range. Meanwhile, an increasing spin localization is observed at the other two carbon atoms as the remaining spin density is distributed among C9 and C11. However, the spin density is the highest for C11 over the whole analyzed dihedral range. Since the curves shown in Figure 5.4 do not cross over the torsion values significantly populated in the active site of the enzyme, it seems unlikely that spin localization is the steering factor for the observed selectivity.

However, the degree of spin localization may sensitively depend on the substrate model employed. This is indicated by comparison of the obtained spin density distribution of the 2,5-heptadienyl radical to the results from Hu and Pratt.^[133] In the corresponding study only a slightly smaller substrate model was employed, the 1,4-pentadienyl radical. Nevertheless, the difference in form of two additional methyl groups at the outer positions of the pentadiene unit already results in changes of the amount of spin localization at the central carbon in a planar model conformation. This indicates that the spin density distribution within the actual fatty acid radical substrate may differ from both of these models due to their limited size. Furthermore, only one scan over a single dihedral angle was performed for the 2,5-heptadienyl system, which does not represent a real conformation of the fatty acid substrate within the enzyme pocket. And finally, none of the small substrate models included the influence of the environment such as the nearby iron center and the surrounding amino acid residues. These elements may have an impact on the resulting spin density distribution of the substrate and it is therefore necessary to analyze the changes in the spin localization concerning the full fatty acid substrate radical under the effect of the closest enzyme environment.

For the CspLOX2 wild type three different sizes of model systems were generated out of snapshots taken from the performed MD simulations. Their spin density distribution was calculated with the local density fitting formalism presented in chapter 4 and with the conventional density fitting approach at the LDF-UB3LYP/def2-SVP level. The involved SCF procedure was iterated until changes in the norm of the resulting electronic density matrix were found to be smaller than $10^{-5} a_0^{-3}$. All three model structures and the corresponding spin density results are shown in Figure 5.5. Only the linoleic acid radical shown in red is covered in the small

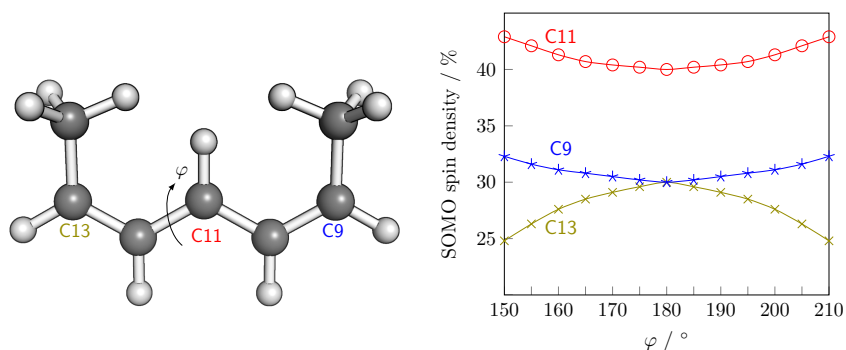


Figure 5.4: Small radical substrate model system 2,5-heptadienyl radical (left). The dihedral angle φ was varied for spin density analysis. Spin distribution of the model system for different torsions of φ (right).

Table 5.1: System sizes and speed ups of the local density fitting approximation compared to the non-local approach for three different model system sizes for the CspLOX2 wild type in UB3LYP/def2-SVP calculations. The number of atoms, number of basis and fitting functions and the average fitting domain size divided by the total number of fitting functions q_{fit} are provided. The average speed up per iteration $\langle f_{\text{iter}} \rangle$ and the total speed up factor f_{tot} of the local density fitting procedure were calculated using 20 cores (Intel 2.8Ghz).

Model	N_{atoms}	N_{bas}	N_{fit}	q_{fit}	$\langle f_{\text{iter}} \rangle$	f_{tot}
small	50	430	1564	0.33	1.10	1.07
medium	167	1545	6035	0.14	5.14	3.28
large	294	2657	10159	0.07	10.79	7.61

model system. A medium-sized model adds the iron moiety and its ligands, namely His257, His262, His449, Ile569, Asn453 and one coordinated water molecule. The large model system further includes the nearby amino acid residues Glu253, Leu258, Ile296, Ala300, Leu304, Leu502, Thr505, and Leu506. Detailed information concerning the system sizes as well as the errors and speed ups of the local density fitting approach is provided in Table 5.1.

In terms of the spin density distribution shown in Figure 5.5 only slight differences are found comparing the local and non-local density fitted calculations for all three system sizes. However, the resulting spin localization

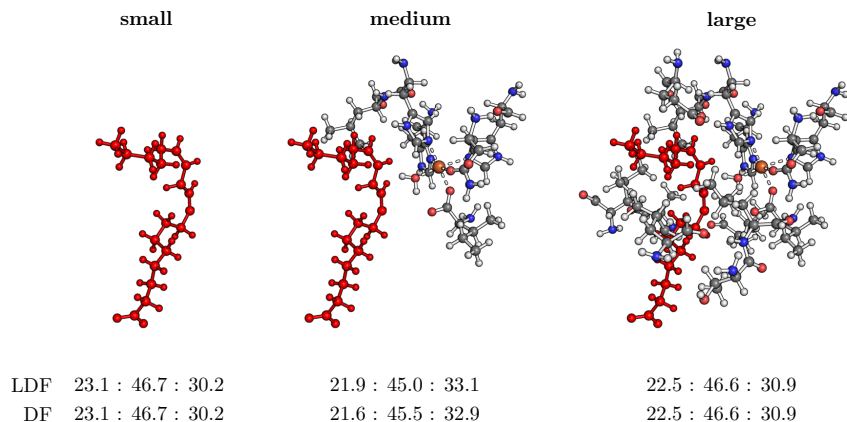


Figure 5.5: Model structures of the CspLOX2 active site around the linoleic acid radical shown in red. The medium-sized model adds the iron moiety and its ligands, namely His257, His262, His449, Ile569, Asn453 and one coordinated water molecule. The large model system further includes the nearby amino acid residues Glu253, Leu258, Ile296, Ala300, Leu304, Leu502, Thr505, and Leu506. Below the structures the corresponding SOMO spin density distribution C9 : C11 : C13 is displayed for the density fitted U3B3LYP/def2-SVP method in their local (LDF) and non-local (DF) variants.

is varying by going from the small to the large model system. While the addition of the iron center in the medium-sized model leads to a clear change of the spin distribution, the effect is partly neutralized with the additional inclusion of the closest amino acid residues in the large model. The error of the local density fitting procedure was calculated for the conformational energy of two different structures of the small model size and was found to be below $0.01 \text{ kJ}\cdot\text{mol}^{-1}$. It is therefore expected that no major errors are introduced to the description of spin localization, which is also indicated by the identical spin distributions of the large model comparing the DF and LDF approaches. While no significant speed ups for the small substrate model of 50 atoms are obtained, a pronounced acceleration of the calculation by application of LDF approximations results for the large model system containing 294 atoms. In this structure the average fitting domain size is only 7% of the total number of density fitting functions, leading to significant speed ups compared to the non-local DF approach. While the total LDF calculation is more than seven times faster than the conventional density fitting, each iteration within the corresponding SCF cycle is even more than ten times as fast as in the non-local procedure. With this significant speed up and the minor errors in the resulting molecular energy and density, the LDF approach provides a perfect tool for the systematic analysis of the spin density distribution within the linoleic substrate radical including the closest enzyme environment.

Structures of the size of the large model system were generated out of the MD snapshots of the CspLOX2 wild type and the L502V single amino acid mutant. These contain the same amino acid residues as shown before with a leucine for position 502 in case of the wild type and a valine for the point mutant yielding in total 294 and 291 atoms, respectively. According

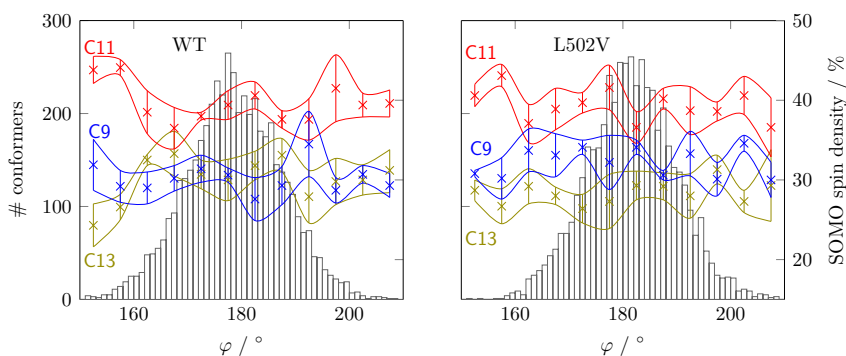


Figure 5.6: Conformer count over the dihedral angle and spin density distribution of the carbons C9, C11 and C13 of the linoleic acid substrate radical for the CspLOX2 wild type (left) and the L502V single amino acid mutant (right). The width of the spin density curves is given by the standard deviation within each dihedral bin.

to their C10-C11-C12-C13 substrate dihedral angle, five structures were chosen for each angular bin of 5° between 150° and 210° . By application of local density fitting approximations as described in section 4.1, the spin density distribution of the substrate was computed employing LDF-UB3LYP/def2-SVP calculations. Convergence of the procedure was assumed, if the norm of the electron density did not change by more than $10^{-5} a_0^{-3}$ compared to the previous SCF step. The spin density information was correlated with the overall dihedral angle distribution in order to respect the biochemical relevance of the corresponding structures. Spin density results for both systems are shown in Figure 5.6.

Looking at the spin distribution of the substrate, the overall results do not drastically change compared to those observed for the small 2,5-heptadienyl model system. For both enzyme species the C11 carbon carries the highest spin density over the whole analyzed dihedral angle window. Even though, no clear trend is found for the out-of-plane bending between C9 and C13, only slight crossing of the corresponding curves with the one from C11 is observed within the standard deviation. The dihedral angle distributions with maxima close to 180° for both enzymes point out additional biochemical relevance to the investigated near-planar substrate structures. If molecular oxygen is assumed to attack the carbon site of highest spin density, a product distribution can be extracted by correlating the spin density results with the dihedral angle distribution.

However, the resulting product ratios are unsatisfactory. Similar product distributions are obtained for both enzymes with C9 : C11 : C13 ratios of 30.7% : 37.4% : 31.9% for the wild type and 32.4% : 38.6% : 29.0% for the L502V mutant. This is in clear conflict with the experimental product distributions of 17.6% : 74.2% : 8.2% for the CspLOX2 wild type and 11.9% : 49.2% : 38.9% for its single amino acid mutant. All in all the hypothesis of spin localization within the substrate would always support C11 as the main site for dioxygen attack. Even though this is the case for the CspLOX2 wild type system it is incapable of explaining the drastically change in product distribution for all of the single amino acid mutants. Taking into account that C11 hydroperoxides are actually not the main product for almost all other LOXs, the spin density distribution of the substrate does not provide a suitable explanation for the observed selectivities.

5.3.2 Activation Barriers

Kinetic aspects of the reaction catalyzed within the LOX class of enzymes have been of particular interest in order to understand the details of the underlying mechanisms and were investigated via the calculation of rate constants and activation barriers. These studies reached from the initial hydrogen abstraction^[118–125] over back and side reactions, such

as β -fragmentation^[141], up to the addition of molecular oxygen on the radical substrate species.^[136] Figure 5.7 displays a schematic profile of the energy following the reaction coordinate from the free fatty acid up to the formed hydroperoxide. While hydrogen atom abstraction is proposed to be the rate-determining step with the highest barrier to cross within the catalytic cycle, the addition of molecular oxygen exhibits different barriers for the attacked carbon centers. The latter is therefore proposed to be the crucial step for the enzymatic selectivity control and is thus analyzed in further detail. Even though it was clearly shown that hydroperoxides at the outer carbons of the pentadiene unit are the thermodynamically favored products^[117], a different picture might occur if one analyzes the kinetic aspects of dioxygen addition.

For the small 2,5-heptadienyl radical model compound, relaxed surface scans at the B3LYP-D3/def2-TZVP level of theory were performed simu-

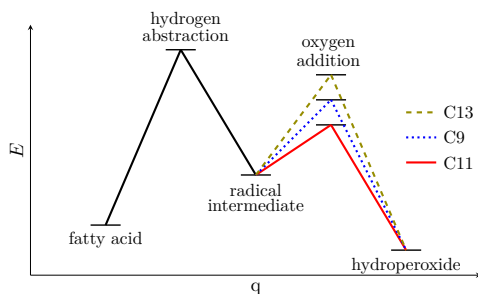


Figure 5.7: Schematic energy profile of the enzyme catalysis in CspLOX2 along the reaction coordinate. The three different pathways for dioxygen addition represent the reaction at C9, C11 and C13.

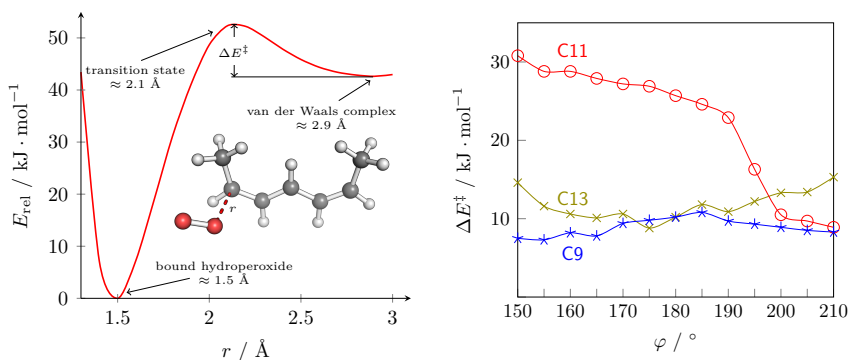


Figure 5.8: Energy profile of a relaxed surface scan for molecular oxygen approaching a reactive carbon site of the small model system (left). Activation barriers for dioxygen addition of different torsions φ obtained by relaxed surface scans (right).

lating the approach of molecular oxygen towards the three carbon sites of interest. Following a path of lowering carbon-oxygen distance, the internal carbon dihedrals corresponding to C9-C10-C11-C12 and C10-C11-C12-C13 in linoleic acid were kept fixed during the scans. The former was fixed at a planar 180° conformation for all calculations while the latter was varied within the 30° window as already applied for the spin density analysis on the small substrate model. Activation barriers of dioxygen attack were extracted from all scans. The curve of a generic surface scan and the activation barrier results within the dihedral angle window are shown in Figure 5.8.

When approaching molecular oxygen towards a reactive carbon site a van der Waals pre-complex is formed at carbon-oxygen distances of typically 2.9 \AA , which can be identified by a local energy minimum along the path. In order to form the bound hydroperoxide the system has to cross the transition state with an energy maximum at about 2.1 \AA . At distances of approximately 1.5 \AA the global energy minimum structure of the hydroperoxide product is found. The electronic activation barrier ΔE^\ddagger of molecular dioxygen attack is defined by the energy difference of the transition state structure and the pre-complex. By looking at the results for conformers close to 180° , a significantly higher activation barrier is found for the attack at the central C11 carbon compared to C9 and C13. In contrast to the spin localization hypothesis this suggests no dioxygen attack to occur at C11 for the planar substrate conformer since the reaction is highly kinetically unfavorable. For this structure the carbon center of maximum spin density is also the center requiring the highest activation barrier for dioxygen addition which is increased compared to the other centers by $15.5 \text{ kJ} \cdot \text{mol}^{-1}$. This difference would rule out any reaction to take place at the central carbon without external influence.

However, slight torsions within the model system allow the reaction paths to become competitive yielding similar activation barriers for dihedral angles around 200° . While this crossing in carbon site preference was not found in the case of the spin density analysis, enzyme-induced substrate bending might lead to significant changes in the activation barriers when dioxygen is approaching. If the conformational distribution of the substrate is significantly different among the analyzed enzyme mutants, changes in the activation barriers could provide a suitable explanation for the strongly varying product distributions. This idea is pursued in the following section.

5.3.3 Substrate Conformation

In order to analyze the relevant substrate conformations the full enzyme structure with the bound linoleic acid radical was simulated in a water box employing molecular dynamics. Starting from the experimental crystal structure of CspLOX2, simulations with three different initial substrate

geometries were performed for the wild type and all single amino acid mutants as described in the computational details. For all 5000 snapshots taken from the total run of 1 ns, the value of the C10-C11-C12-C13 dihedral was extracted for each simulation and combined in a histogram of conformers in a range of 60° . The results for the CspLOX2 wild type and the five point mutants are shown in Figure 5.9.

Even though the distributions differ in shape and location of their maximum peaks, the average dihedral angle position is close to the preferred planar substrate conformation reaching from 176.9° in the case of the

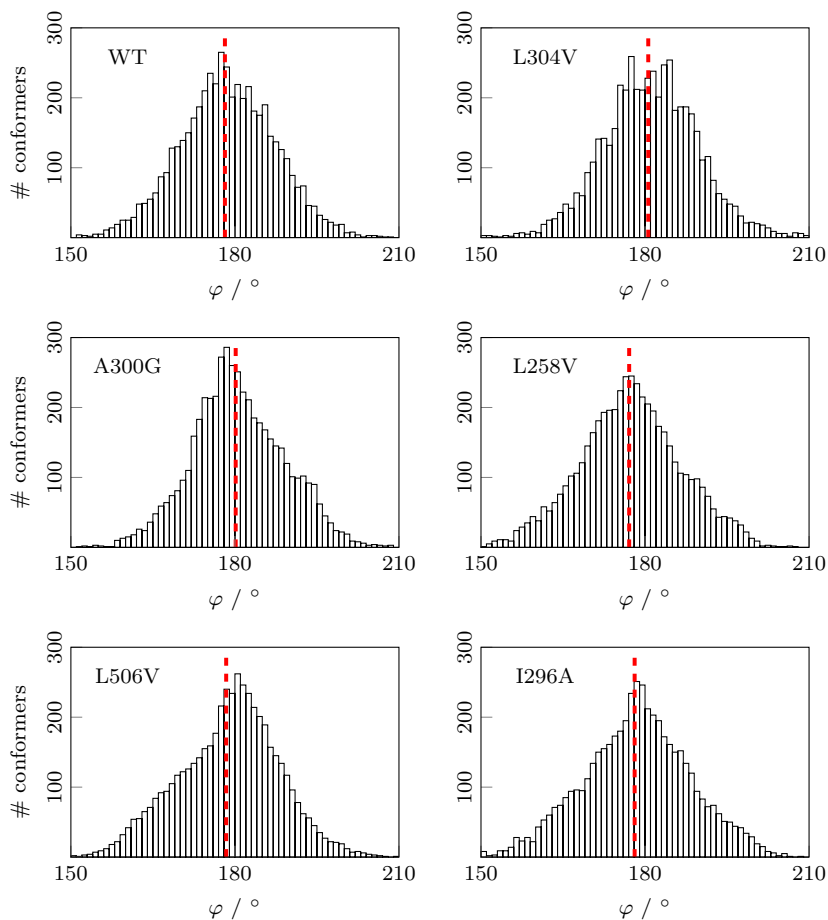


Figure 5.9: Histograms of conformers according to the C10-C11-C12-C13 dihedral of the linoleic acid substrate from MD simulations on the CspLOX2 wild type (top left) and the single amino acid mutants L304V, A300G, L258V, L506V and I296A. The average value of each distribution is indicated by a dashed red line.

L258V mutant over 178.2° for the wild type up to 180.6° for the L304V mutant. A similar behavior was already seen for the L502V mutant in Figure 5.6 with an average dihedral angle of 181.1° . For all simulations the substrate seems to be flexible inside the enzyme pocket, allowing for internal torsions in the analyzed dihedral angle window. However, no significant difference in the preferred conformation is observed among all analyzed enzymes. The activation barrier calculations for the small model system in the previous section showed that the dihedral angle would have to change by about 20° compared to the planar conformation to allow the reaction paths towards the different carbon sites to become competitive. If changes in the activation barriers due to substrate torsion are assumed to be the steering factor for the enzyme selectivity, nothing more but similar product distributions would be found on the basis of the analyzed histograms. With similar substrate structures of relevance for all analyzed enzymes, but their strongly deviating experimental product distributions, enzyme-induced substrate torsions modifying the activation barriers do not offer a plausible selectivity mechanism. Other influences within the enzyme environment guiding the reaction towards a clear pathway need to be responsible for the observed selectivity. Such guiding might include the existence of an oxygen channel, transporting dioxygen directly towards specific carbon atoms or the sterical blockage of other reactive carbon sites forcing the addition at the available position. These possibilities are further discussed in the following section.

5.3.4 Steric Shielding

Enzyme-induced substrate conformations leading to changes of the substrate spin density or in the activation barriers for dioxygen attack could be effectively ruled out as relevant factors for the observed enzyme selectivity as seen in the last sections. As such, the possibility of sterical constraints within the active site deserves detailed investigation. The existence of a specific oxygen channel is a widely held hypothesis^[128;135;136;142], but for most LOXs no clear path for dioxygen access could be found. Indeed, molecular oxygen might even enter the enzyme pocket through the same channel as the fatty acid substrate does. Another possible sterical constraint consists in the mechanical blockage of reactive substrate sites by the enzyme. Models for the analysis of sterical shielding around the substrate have been already proposed and applied before.^[135;136] Furthermore, they are normally based solely on the smallest distance between potentially shielding atoms and the substrate carbon sites of interest. Even though this type of model provides the possibility to qualitatively estimate the average space around the reactive carbons, it is incapable of predicting the enzyme product distribution in a quantitative manner. In addition, the shielding results of these models need to be treated with caution as they do not include angular

information. If an atom is close to a reactive carbon site but strongly bent compared to the path in which molecular oxygen approaches, it would lower the average free space around the site even though no particular shielding is present at all. This is the main reason for the limited predictive power of these established steric shielding models.

This work approaches the problem of steric shielding within the enzyme pocket in a different way. Instead of investigating whether the space close to a carbon site is sufficient for the bound hydroperoxide structure, it is assumed that the van der Waals pre-complex needs to be formed first. Thus, in order for the reaction to happen, the corresponding space is required close to a reactive center. With the implicit assumption of an oxygen channel only allowing an attack from the antarafacial substrate side opposed to the iron center, a new steric shielding model was developed. For each analyzed structure an equalization plane through the substrate pentadiene unit was defined by three points, using the centers of mass of C9-C10-C11, C10-C11-C12 and C11-C12-C13 respectively. These points ensures that the constructed plane takes into account the impact of all carbons of the pentadiene unit without the problem of near-collinearity which would rise for C9, C11 and C13 in a zigzag chain configuration. A point in space was found following the normal of the plane with a distance of 3 Å for the respective carbon C9, C11 or C13. This distance corresponds to the van der Waals minimum found in the relaxed surface scans for approaching dioxygen to the small model system as seen in section 5.3.2. The size of molecular oxygen was calculated from a bond distance of 1.21 Å added with half of the van der Waals radius of a single oxygen atom of 1.52 Å. This leads to an additional space needed of about 2 Å relative to the defined point in space in order to fit dioxygen. For each structure the distance r of the enzyme or substrate atom closest to the aforementioned point was found and the relative accessibility $A(r)$ of the corresponding carbon site was calculated as

$$A(r) = \begin{cases} 1 & \text{if } r \geq r_{\max} \\ \frac{1}{r_{\max} - r_{\min}} (r - r_{\min}) \exp(r - r_{\max}) & \text{if } r_{\min} < r < r_{\max} \\ 0 & \text{if } r \leq r_{\min} \end{cases} \quad (5.1)$$

If there was no atom found closer than $r_{\max} = 2.2$ Å, the corresponding carbon site was considered to be fully accessible for dioxygen attack. If an atom was found closer than $r_{\min} = 1.8$ Å, the carbon site was counted as completely shielded as the space is insufficient to form the pre-complex. An exponential damping was used for intermediate distances in order to account for weak steric repulsion and avoid a step function which would lead to large accessibility fluctuations. This steric shielding model is illustrated in the snapshot shown in Figure 5.10. The spheres on the antarafacial

side of the blue substrate are either free for dioxygen attack or their accessibility is reduced by surrounding residues or substrate atoms shown in red. Due to the used damping function movements of the amino acid residues or the substrate are allowed without complete changes in the calculated accessibilities. The sphere model also accounts for the fact that dioxygen usually approaches in a tilted trajectory compared to the normal of the substrate plane. A segment of the Perl script carrying out the steric shielding analysis can be found in the end of the Appendix.

The steric shielding model described above was applied on all 5000 snapshots of each molecular dynamics simulation of the CspLOX2 wild type and its single amino acid mutants. Since the model is only applicable for an antarafacial dioxygen attack, these mutants were the ones conserving the stereoselectivity compared to the wild type yielding C9-*R*, C11-*R* and C13-*S* products. Calculating the accessibility for dioxygen attack on the three carbon centers C9, C11 and C13 resulted in a ratio of available sites over all structures for each simulation. The ratio was used to analyze the product distribution of the corresponding enzyme. An additional penalty taking 10% intensity of C11-*R* and distributing it equally to C9-*R* and C13-*S* was applied in order to account for the in general higher activation barrier at the central carbon moiety. The ratios of the three simulations for each enzyme were averaged and the standard deviation with an additional 5% was used as a basis for an error estimation. This allowed for the

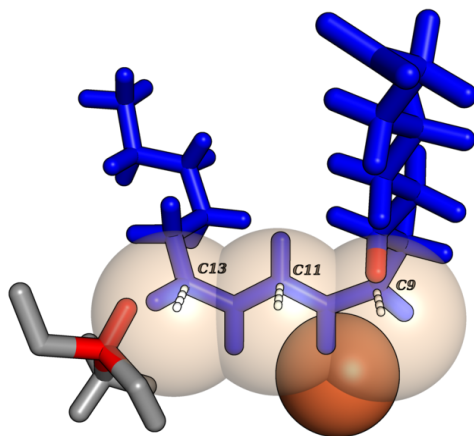


Figure 5.10: Snapshot of the linoleic acid substrate (blue) considered for the steric shielding model. If atoms (red) are found within a sphere close to the carbon atom of interest, the corresponding site is considered to be less accessible. In the case shown, C9 would be partly shielded by a substrate atom, while the accessibility of C13 is reduced by an amino acid residue. C11 would be the only site to be considered fully accessible. The iron of the active site is shown in the background.

comparison of the simulated product distributions with the ones observed by experiments for an antarafacial dioxygen attack as shown in Figure 5.11.

The agreement between the experimental and calculated product distributions for the considered enzymes is remarkable. With a maximum deviation of 7.4% over all products and a root-mean-squared deviation of only 3.1% the calculations are capable of accurately reproducing the experimental enzyme selectivity. To the best of my knowledge this is the first theoretical model to quantitatively replicate the specificity of several LOXs mutants. The agreement with the experimental distribution is a clear evidence for steric shielding and blockage of reactive carbon sites considerably determining the enzyme selectivity. With controlled amino acid residue movements, the space around specific carbon centers is restricted by the enzyme, preventing molecular oxygen to form the van der Waals pre-complex. Since the pre-complex is essential to cross the transition state towards hydroperoxide formation, the corresponding carbon sites are effectively shielded and no reaction takes place at these centers. This drives the oxygenation towards non-shielded carbon atoms and yields the observed specificity for each enzyme.

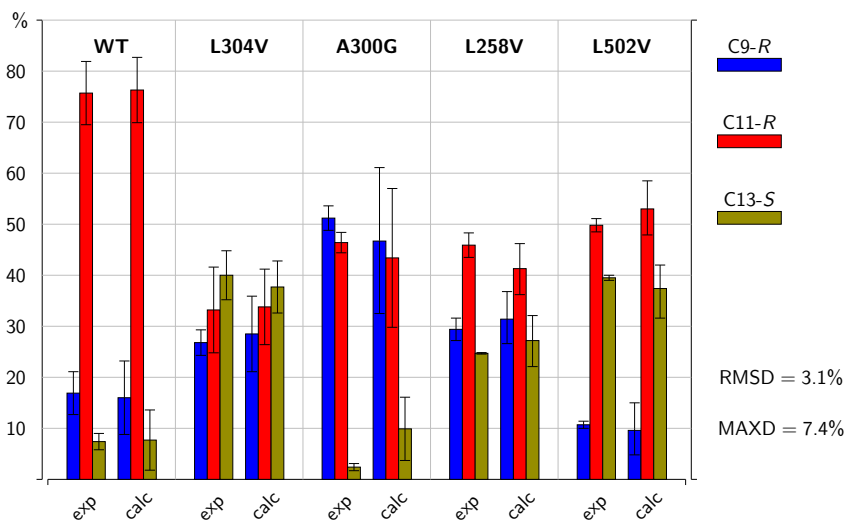


Figure 5.11: Product distributions for an antarafacial dioxygen attack in CspLOX2 and investigated mutations. The distributions of the main stereo products of the CspLOX2 wild type (WT) and a set of single amino acid mutations were calculated from MD simulations using a steric shielding model (calc) and are compared to the experimental distributions (exp). A penalty was applied to C11-R, reducing its intensity by 10% and equally distributing it to C9-R and C13-S. The error bars correspond to the standard deviation with an additional 5% for the calculated distributions to respect the approximations involved.

The existence of an oxygen channel was by no means excluded using the developed shielding model. It might still be the case that the enzyme transports molecular oxygen through a specific channel towards the iron opposed substrate side to enable an antarafacial dioxygen attack. Considering only enzymes that preserve the stereochemistry of the wild type within this model, the inversion of single stereo products for two of the single amino acid mutants is still unexplained. Due to the remarkable performance of the steric shielding model among the enantioselectivity conserving enzymes, it seems unlikely that a molecular oxygen approaches via a different channel from the suprafacial substrate side, yielding inverted products at C9 and C13 for the I296A and L506V mutants. Most likely, the mutated enzymes enforce a restricted substrate binding with the affected carbon moiety being flipped. This hypothesis is strengthened by the fact that the mutated isoleucine residue 296 is close in space to the C9 substrate moiety, while leucine 506 is found in the vicinity of C13 in the wild type structure. Since both mutations convert a larger amino acid into a smaller one, more space is provided at the corresponding region on the side of the iron center, allowing for a one-sided substrate twist. The consecutive antarafacial dioxygen attack would then yield inverted stereo products at the carbon sites involved. However, no evidence for the existence of such structures could be found via the application of molecular dynamics simulations with a manually flipped substrate. The used force field parameters for the radical pentadiene unit do not allow for this conformation due to the limited parametrization, even though it might be the preferred geometry in the I296A and L506V enzymes.

5.4 Conclusion

The cyanobacterial lipoxygenase *CspLOX2* covers a rare case of the bisallylic C11 hydroperoxide being the clear main product in amounts of more than 70%. Single amino acid mutations in the vicinity of the active site of the enzyme are capable of drastically changing the product distributions leading to significantly lower amounts of C11 product and in some cases even to the preference of a conjugated hydroperoxide at C9 or C13. This indicates that the *CspLOX2* wild type possesses a specific mechanism to control the oxygenation towards the production of the bisallylic product. Several factors proposed to be responsible for the enzyme selectivity were investigated by using a variety of computational approaches.

The assumption that the carbon site of highest spin density is preferred for molecular dioxygen attack would deliver a simple explanation for the observed specificity due to enzyme-induced spin density localization of the substrate. DFT calculations on a small model systems showed that changes in the substrate conformation by twisting the pentadiene unit out of the planar arrangement are indeed varying the spin density distribution, but

not in a way that the site of highest spin localization would differ. Even DFT calculations on a larger model system including the iron center and nearest amino acid environment did not change this picture. Considering the biological weight of the substrate conformers, the central C11 carbon was always the center of highest spin density. Even though this approach might therefore favor the production of the bisallylic hydroperoxide as observed in the CspLOX2 wild type, it can not explain the conjugated main products in other LOXs and the CspLOX2 mutants investigated by the Feussner group.^[132] The hypothesis of a spin-guided dioxygen attack would suggest the reaction to preferably occur at this central position and the bisallylic hydroperoxide should be the main products for all LOXs. Since this is in clear conflict with the experimental findings, spin localization within the substrate was effectively ruled out as an explanation for the enzyme selectivity.

Another specificity-controlling mechanism was investigated in form of changes in the activation barriers for molecular oxygen attack. Relaxed surface scans approaching dioxygen towards the reactive carbon sites of a small model system gave a clear contrast to the spin-guided hypothesis, as the activation barrier was highest for the preferred planar substrate conformation. The activation barriers for other substrate conformers indicated that slight substrate torsions are sufficient to let the different reaction pathways become competitive. Molecular dynamics simulations were carried out to investigate the preferred substrate conformations within the CspLOX2 wild type and its single amino acid mutants. Different degrees of substrate twisting when bound to the active site of the enzyme would provide an explanation for the observed selectivities as they go hand in hand with changes of the activation barrier for dioxygen attack. Nevertheless, no significant changes in the relevant substrate dihedral angle were found in the different enzyme simulations. The induction of varying substrate geometries by the enzyme and with it the changes in the activation barriers for dioxygen attack were therefore excluded as steering factors for product specificity.

The possibility of an oxygen transport channel delivering molecular oxygen towards specific carbon sites as well as sterical blockage of carbon sites by amino acid residues of the enzyme were investigated by using molecular dynamics simulations. A simple shielding model regarding only free space information around reactive carbon atoms was used to analyze the accessibility of molecular oxygen to form the corresponding van der Waals pre-complex. With an additional penalty respecting the higher activation barrier at C11, the resulting ratios of accessible carbon sites were capable of accurately reproducing experimental product distributions. This is a clear indication that steric shielding of specific carbon atoms is the main selectivity mechanism present in the CspLOX2 enzyme. Since the steric shielding model only considered antarafacial dioxygen attacks

and still yielded remarkable accuracy compared to the experimental data, the existence of an oxygen channel transporting molecular oxygen to the antarafacial substrate side, opposed to the iron moiety, is supported. The developed model for steric shielding analysis was the first to accurately reproduce experimental product distributions in LOX systems.

6 Summary

The aim of this work was the development and application of low-scaling self-consistent field approaches employing local and fragment potentials. This was of particular interest for the Hartree-Fock method, since its computation was found to possibly act as a computational bottleneck in single-reference local correlation treatments.^[1-3;5;6] A central task was therefore the search for viable approximations to accelerate the calculation of the Hartree-Fock solution. With this purpose in mind, two different approaches have been followed in the course of this thesis.

In a first approach to the problem a cap-free fragmentation was employed to form the proposed FJK method.^[66] Within the FJK procedure, the molecular system under consideration was split up into smaller entities, which were subsequently separately treated. Recovery of the complete Hartree-Fock description was obtained by an iterative process, very much in the spirit of the self-consistent field approach itself. By inclusion of full Fock potentials of the remaining system in all fragment calculations, each entity was embedded in a mean-field. Iteration over the generated fragments allowed for the convergence of the FJK results to the full Hartree-Fock solution of the investigated system. This property was extensively analyzed in terms of the constructed molecular density as well as the absolute and relative energy results for a set of molecular benchmark structures. No particular issues regarding the convergence of the FJK approach were obtained for the allowed generation of charged and non-physical fragments. Even in the case of a completely delocalized system, the correct Hartree-Fock solution could be restored within small energy errors by an increasing number of fragment iteration cycles.

In the original formulation of the FJK procedure a full molecular basis set was employed for each fragment calculation to cancel possible basis set superposition effects. The high computational effort arising from this treatment ruled out a favorable scaling of the method in the initial form. Even though it was shown that the number of SCF iteration steps within each fragment calculation could be restricted without introduction of major errors, the computational effort is not sufficiently reduced by such an approach. To overcome this problem, a modified FJK treatment was proposed employing the large basis set only for the calculated fragment, while a minimal basis set representation was introduced for the embedding potentials. Although the possibility to converge to the full Hartree-Fock results was widely kept by this procedure, it could be shown that the involved fragment calculations allow for a favorable scaling at best in a very narrow window of system size only. Thus, no fragment-based procedure of overall reduced computational cost compared to the conventional Hartree-Fock method could be established. This would be only possible in the limit of size-independent fragment calculations, which may be obtained

by truncation of the embedding area and restriction of the included Fock potentials to fragments of the closest environment.

Nevertheless, the application of full Fock potentials provided an exact and stable approach for the representation of a molecular environment. The procedure was employed as a reference to analyze the embedding potentials of QM/MM simulations in condensed phase. Investigation of the resulting molecular density under the embedding effect of nearby monomers pointed out that significant differences compared to the inclusion of full Fock potentials are possible in the application of simple point charge models. Separation of the Fock contributions into Coulomb and exchange effects showed the additional influence of embedding exchange potentials to be comparably small. The major effect was arising from the Coulomb contributions, which could not be recovered by atomic point charges for monomers close to the calculated molecule. However, the QM/MM potentials were found to be sufficient to describe the long-range impact of further distant monomers on the resulting molecular density. Thus, the study confirmed the concept followed by multi-layer descriptions, in which the closest environment is represented by sophisticated potentials, while the distant interactions are included at a more approximate level of embedding.

The representation of a molecular environment by an exact embedding potential provides the starting point for other possible applications. In periodic structures like molecular crystals it would be possible to employ the embedding with full Fock potentials on the basis of a single monomer calculation and subsequent projection of the electron density to the other embedding molecules. The embedding procedure could then be performed in an iterative procedure to converge the description of the crystal structure as already proposed for QM/MM methods.^[77] Another application of quantum mechanical approaches with an embedding potential is the calculation of phenomena in heterogeneous catalysis by combination of wave function methods and embedded cluster models.^[143;144] In such calculations the full Fock embedding could improve the description of the critical interactions between substrate and catalyst and thus provide insights into the underlying catalytic mechanism.

An entirely different approach to the problem of the steep scaling of computational cost in Hartree-Fock theory followed the idea to directly address the computationally most demanding step, the evaluation of exchange contributions. By employment of a localized molecular orbital basis the rapid decay of exchange interactions with an increasing interatomic distance was exploited and a linear scaling algorithm for the computation of the exchange matrix resulted. The corresponding local density fitting (LDF) approach was already established and implemented for closed-shell Hartree-Fock calculations.^[9] This moved the focus to the application of

the method to open-shell radical structures. Besides the restricted and unrestricted formulations of the Hartree-Fock procedure, the LDF approximations were also applied to hybrid density functional approaches, making use of the algorithm in the admixed contributions of exact Hartree-Fock exchange. Different type of benchmark sets were employed to investigate the limitations and the speed up gains of the local fitting of the exchange matrix. Strongly delocalized systems provided insights into the energy errors of the procedure in extreme cases. With an increase of the π -character of the investigated structure, the LDF approximations became less accurate as larger errors in the total energy description were found. However, the energy difference to the non-local variant of the procedure was below the intrinsic error of the density fitting approach itself even for the most complicated case of the study. The corresponding structure was employed to investigate the possibility of error control in terms of the introduced thresholds for the fitting approximations. Indeed, by adjusting the thresholds of sparsity of the localized molecular orbital coefficients and the fitting domain size, the results were found to smoothly converge to the conventional density fitting descriptions.

Another benchmark set consisted of RAFT initiation reactions which provided a series of radical structures of different molecular size. The application of LDF approximations in restricted and unrestricted formulations of Hartree-Fock and hybrid DFT approaches yielded access to the analysis of the errors in relative energies and the speed ups compared to the non-local calculations. Furthermore, the benchmark set allowed for the determination of the critical molecular size, upon which a favorable scaling results from the applied LDF procedure. While for all systems considered no significant errors in the relative electronic reaction energies were obtained, the resulting speed ups varied among the radical structures. The acceleration obtained by the local fitting of exchange increased with the molecular system size. A comparable trend could be followed for an expansion of the applied basis set within each benchmark structure. For small and medium basis sets the critical system size of favorable scaling could be identified to be located in a region between 36 and 50 atoms. For larger radical structures the speed up gains of the local density fitting were more pronounced reducing the computational effort compared to the non-local variant by a factor of 2 and higher. This favorable scaling and the possibility to control the resulting error via the adjustable fitting thresholds allows for the accurate calculation of large structures within heavily reduced computational timings. Thus, the LDF approach enables the desired properties of a low-scaling self-consistent field procedure and can provide access to manifold applications in macromolecular and biomolecular studies.

Selectivity analysis of the cyanobacterial lipoyxygenase CspLOX2 provided a problem of high complexity. In the dioxygenation reaction of a

polyunsaturated fatty acid the CspLOX2 enzyme covers the rare case of bisallylic product formation with an unclear mechanism involved. Several computational models and methods including the presented local density fitting approach were applied to investigate the different proposed factors steering the product selectivity. A small radical substrate model was employed to investigate the hypothesis of spin localization within the pentadiene unit of the fatty acid radical leading to preferred dioxygen attack at spin-enriched carbon centers. The results indicated the changes in spin density distribution to be insufficient for control of the product formation. However, the description of this small isolated molecule to represent the substrate within the enzyme pocket was somewhat incomplete. The lacking effects of the iron moiety and the surrounding amino acid residues were taken into account by generation of a large model system of the active site. Local density fitting approximations were employed in hybrid DFT calculations to complete the analysis of spin density within the fatty acid radical. Indeed, the results confirmed the indications of the small model, according to which spin localization could be ruled out as suitable factor of enzyme selectivity.

Relaxed surface scans with molecular oxygen approaching to the small substrate model provided access to the activation barriers of dioxygen attack at the different carbon centers. In contrast to the results of the spin density analysis, it was found that slight substrate torsions are sufficient to allow for competitive reaction pathways. This idea was followed by molecular dynamics simulations of the entire enzyme structure, in order to sample the conformational states of the substrate bound to the active site. However, no differences in the preferred structure of the substrate radical could be identified among the CspLOX2 wild type and the investigated set of its point mutants. Indeed, the simulations assigned particular biochemical relevance to structures with the involved pentadiene unit in a nearly planar conformation, in which no reaction at the central bisallylic carbon atom would occur due to the high activation barrier for oxygen attack. This moved the focus of the enzyme selectivity analysis to the last remaining factor proposed, the shielding of specific carbon sites in terms of steric blockage.

A new model was developed to analyze the effect of steric shielding within the lipoxygenase enzyme structure. On the basis of an equalization plane through the substrate, a point in space in front of the reactive carbon site was correlated to the formation of a van der Waals pre-complex in the dioxygenation addition. The corresponding structure was identified within the relaxed surface scans of the small substrate model. It was proposed that the corresponding carbon atom is accessible for molecular oxygen addition, if the space around the defined point is sufficient to fit the dioxygen molecule and thus form the pre-complex, which is mandatory for product formation. After application of an additional penalty to respect the

differences in the activation barriers, the ratio of accessible reactive carbon sites within the snapshots taken from molecular dynamics simulations could be directly compared to experimental product distributions. This yielded a remarkable agreement with the experimental product amounts for the CspLOX2 wild type and its single amino acid mutants. Since the proposed shielding model is solely based on the geometrical constitution around the pentadiene unit of the substrate, the effect of steric blockage could be identified as the main factor to the resulting enzyme selectivity.

A comprehensive picture of the CspLOX2 enzyme structure and reactivity has been recently published.^[132] The presented theoretical study of the enzyme and its mutants provided supplement to the experimental data and yielded detailed insights into the underlying selectivity mechanism. Former studies could already correlate the amount of steric shielding with the resulting product distributions in soybean lipoxygenase^[135] and cyclooxygenase^[130;136] structures, but were limited to a qualitative analysis. The here proposed steric shielding model is the first in literature to accurately reproduce experimental product distributions in lipoxygenase enzyme structures. Its application to the abovementioned structures could therefore conclusively clarify the impact of sterical blockage within the corresponding enzymes in a quantitative manner.

Appendix

This appendix contains the structures that are used in the context of this work. On the end a small segment of the Perl script performing the steric shielding analysis is provided. All geometries are either given as a Z-matrix with distances in Å and angles and dihedrals in degrees or in the xyz-format in Å.

A.1 FJK Benchmark Structures

The value of the dihedral angle `torsion` was varied from -180° to 180° in steps of 10° to generate 36 conformers of each structure. The carbon-carbon bond length `ccdist` changes values from 1.21 Å to 1.35 Å in steps of 0.01 Å yielding 14 structures in the case of benzene.

Propane

```
C
C 1 1.522920
C 1 1.522920 2 111.937
H 1 1.090900 2 109.584 3 -121.792 0
H 1 1.090900 3 109.584 4 121.792 0
H 3 1.089990 1 110.739 2 torsion 0
H 3 1.089990 1 110.739 6 122.002 0
H 3 1.089990 1 110.739 6 -122.002 0
H 2 1.089990 1 110.739 3 180.000 0
H 2 1.089990 1 110.739 9 122.002 0
H 2 1.089990 1 110.739 9 -122.002 0
```

Butane

```
C
C 1 1.52282164
H 2 1.08984935 1 110.75975859
H 2 1.08984043 1 110.73920579 3 119.22506203 0
H 2 1.08871157 1 111.79836024 4 120.32473226 0
H 1 1.09236168 2 109.78094223 3 180.00000000 0
C 1 1.52315630 2 112.53778388 6 121.52823311 0
H 1 1.09229539 2 109.84823056 6 243.26904112 0
C 7 1.52280262 1 112.57729782 2 torsion 0
H 7 1.09237474 1 109.05618901 9 122.11129158 0
H 7 1.09234070 1 109.07813340 9 237.87345850 0
H 9 1.08875772 7 111.74886510 1 180.00000000 0
H 9 1.08981992 7 110.77861326 12 120.34244808 0
H 9 1.08982181 7 110.77041179 12 239.70975832 0
```

Pentane

C					
C	1	1.52279474			
H	2	1.09228842	1	109.12041946	
H	2	1.09229385	1	109.10399495	3 115.83501208 0
C	2	1.52305937	1	112.55100840	4 122.06703299 0
H	1	1.09395312	2	109.25155141	3 180.00000000 0
C	1	1.52283039	2	113.27302721	6 121.99822002 0
H	1	1.09391896	2	109.28094890	6 244.08678334 0
C	7	1.52308313	1	112.54554275	2 torsion 0
H	7	1.09229096	1	109.09465558	9 122.04873048 0
H	7	1.09226666	1	109.13539229	9 237.89820750 0
H	9	1.08881805	7	111.74038355	1 180.00000000 0
H	9	1.08985070	7	110.78072074	12 120.33673597 0
H	9	1.08982807	7	110.76512023	12 239.70423779 0
H	5	1.08881770	2	111.74037753	1 180.00000000 0
H	5	1.08984139	2	110.76638037	15 120.30651684 0
H	5	1.08983554	2	110.78088385	15 239.67472190 0

Hexane

C					
C	1	1.52307578			
H	2	1.08988719	1	110.75716726	
H	2	1.08988864	1	110.74274288	3 119.23023274 0
H	2	1.08879584	1	111.79147161	4 120.32608708 0
H	1	1.09227398	2	109.74861288	3 180.00000000 0
C	1	1.52314355	2	112.50258044	6 121.57992615 0
H	1	1.09223795	2	109.79430021	6 243.30280907 0
C	7	1.52252274	1	113.24398661	2 180.00000000 0
H	7	1.09383652	1	109.22844732	9 122.08597463 0
H	7	1.09383600	1	109.21557768	9 237.95726873 0
C	9	1.52304385	7	113.24608933	1 torsion 0
H	9	1.09382984	7	109.30568402	12 122.01560878 0
H	9	1.09383943	7	109.31654097	12 237.98609759 0
C	12	1.52306149	9	112.53860532	7 180.00000000 0
H	12	1.09228385	9	109.11626397	15 122.07881168 0
H	12	1.09229498	9	109.11611557	15 237.92454248 0
H	15	1.08884037	12	111.74216741	9 180.00000000 0
H	15	1.08986240	12	110.77388537	18 120.31681448 0
H	15	1.08986937	12	110.77461281	18 239.68444924 0

Benzene

C					
H	1	1.08142068			
C	1	ccdist	2	120.00001756	
C	3	ccdist	1	120.00001756	2 180.00000000 0
C	4	ccdist	3	120.00001756	1 0.00000000 0
C	5	ccdist	4	120.00001756	3 0.00000000 0
C	6	ccdist	5	120.00001756	4 0.00000000 0
H	3	1.08142068	1	120.00001756	4 180.00000000 0
H	4	1.08142068	3	120.00001756	5 180.00000000 0
H	5	1.08142068	4	120.00001756	6 180.00000000 0
H	6	1.08142068	5	120.00001756	4 180.00000000 0
H	7	1.08142068	6	120.00001756	5 180.00000000 0

Benzophenone

C					
C	1	1.405960			
C	2	1.407008	1	119.000	
C	3	1.394016	2	120.579	1
C	4	1.400643	3	120.010	2
C	5	1.397943	4	119.907	3
C	2	1.505041	3	117.575	4
O	7	1.219999	2	119.660	3
C	7	1.504740	2	120.639	3
C	9	1.405967	7	123.231	2
C	10	1.398152	9	120.360	7
C	11	1.397916	10	120.128	9
C	12	1.400644	11	119.900	10
C	13	1.394060	12	120.022	11
H	10	1.091729	9	120.063	14
H	11	1.093195	10	119.783	9
H	12	1.093455	11	120.031	10
H	13	1.093272	12	120.028	11
H	14	1.091994	9	118.348	10
H	3	1.092009	2	118.271	1
H	4	1.093261	3	119.949	2
H	5	1.093447	4	120.052	3
H	6	1.093260	5	120.086	4
H	1	1.091763	2	120.075	3

Alanine dipeptide

N					
C	1	1.46942			
C	2	1.53680	1	109.7324	
N	3	1.33663	2	116.5159	1
O	3	1.22748	2	121.0285	4
H	1	1.01295	2	117.4644	3
C	2	1.53776	1	109.5284	3
H	2	1.09189	1	109.4164	3
H	7	1.09098	2	109.7353	1
H	7	1.09069	2	109.5211	9
H	7	1.09022	2	110.6541	9
C	1	1.33104	2	124.5789	6
C	4	1.45966	3	121.5433	2
C	13	1.54433	4	109.4865	3
O	14	1.23475	13	120.1538	4
N	14	1.33842	13	117.2575	15
H	4	1.00654	3	116.8378	13
C	13	1.53419	4	108.5194	14
H	13	1.09174	4	109.5160	14
H	18	1.09098	13	109.7668	4
H	18	1.09049	13	109.8419	20
H	18	1.09121	13	110.0861	20
O	12	1.22338	1	123.7235	2
C	12	1.51056	1	116.1613	23
H	24	1.08790	12	109.2866	1
H	24	1.08878	12	109.2772	25
H	24	1.08905	12	108.9498	25
C	16	1.46183	14	123.9658	13
H	16	1.00806	14	117.9663	28
H	28	1.09204	16	109.6765	14
H	28	1.09093	16	110.3255	30
H	28	1.09196	16	109.8176	30

A.2 Embedding Analysis Structures

For the water dimer the value of `ddim` was varied from 1.50 Å to 3.50 Å in steps of 0.1 Å to obtain the structures used for the creation of Figure 3.6. In the QM/MM structures the atoms of the calculated unit are colored in red, while the remaining structure was used for embedding.

Water dimer

```
O 1
H 1 0.942
H 1 0.948 2 105.937
O 3 ddim 1 171.024 2 172.7 0
H 4 0.944 3 107.134 1 63.4 0
H 4 0.944 3 106.787 1 309.9 0
```

CO₂ crystal

```
33
C -0.000014 -0.000014 -0.000014
O 0.675400 0.675403 0.675406
O -0.675386 -0.675389 -0.675391
C -2.646125 -0.000015 2.645965
C -3.321546 -0.675433 3.321380
O -1.970748 0.675359 1.970593
C 0.083784 2.729829 2.729733
C 0 -0.591637 3.405246 2.054319
O 0.759161 2.054454 3.405105
C 2.729902 0.083776 2.729713
O 2.054482 -0.591642 3.405128
O 3.405279 0.759150 2.054341
C -0.000007 -2.646126 2.646002
O -0.675427 -1.970708 1.970587
O 0.675370 -3.321500 3.321374
C -2.646133 2.646096 0.000006
O -1.970712 1.970679 -0.675409
O -3.321510 3.321470 0.675378
C 2.729895 2.729887 0.083754
O 3.405315 2.054470 -0.591661
O 2.054518 3.405261 0.759126
C 2.646104 -2.646067 0.000022
O 3.321524 -3.321485 -0.675392
O 1.970727 -1.970693 0.675394
C -2.729924 -2.729858 -0.083725
O -2.054503 -3.405276 -0.759140
O -3.405301 -2.054484 0.591647
C -2.729873 -0.083747 -2.729741
O -3.405294 -0.759164 -2.054327
O -2.054496 0.591627 -3.405113
C 0.000036 2.646097 -2.645974
O -0.675385 3.321514 -3.321388
O 0.675413 1.970723 -1.970601
C 2.646154 0.000044 -2.645994
O 1.970734 -0.675373 -1.970579
O 3.321531 0.675418 -3.321366
C -0.083755 -2.729857 -2.729705
O -0.759175 -2.054440 -3.405119
O 0.591622 -3.405232 -2.054333
```

Cyanamide crystal

```
50
N 0.081468 0.338558 -0.455389
N 0.009514 -0.968371 1.671975
C 0.017478 -0.368052 0.661043
H 0.485826 -0.088320 -1.288184
H -0.594285 1.086185 -0.589446
N 1.435318 3.819962 -0.577742
N 1.507273 2.513034 1.549622
C 1.499308 3.113352 0.538690
H 1.030960 3.393084 -1.410537
H 2.111071 4.567590 -0.711799
N 1.407215 -3.126805 -0.386825
N 1.479169 -4.433734 1.740539
C 1.471205 -3.833415 0.729607
H 1.002857 -3.553683 -1.219621
H 2.082968 -2.379178 -0.520882
N -3.261691 0.270953 2.117692
N -3.333645 -1.035975 -0.009672
C -3.325681 -0.435657 1.001260
H -2.857333 -0.155925 2.950488
H -3.937444 1.018581 2.251749
N 3.366460 0.306331 1.999066
N 3.294506 -1.000598 -0.128297
C 3.302470 -0.400280 0.882634
H 3.770818 -0.120548 2.831862
H 2.690708 1.053958 2.133123
N -1.850420 -2.050051 -2.197938
N -1.778465 -0.743122 -4.325302
C -1.786429 -1.343441 -3.314370
H -2.254778 -1.623173 -1.365142
H -1.174667 -2.797678 -2.063881
N -1.907841 3.752358 1.995339
N -1.835886 2.445429 -0.132025
C -1.843850 3.045748 0.878907
H -2.312199 3.325480 2.828135
H -1.232088 4.499985 2.129396
N -1.935944 -3.194410 2.186255
N -1.863989 -4.501338 0.058891
C -1.871954 -3.910210 1.069823
H -2.340302 -3.621288 3.019051
H -1.260191 -2.446782 2.320312
N 1.335336 -2.252601 3.904675
N 1.407290 -0.945672 6.032039
C 1.399326 -1.545991 5.021107
H 0.930978 -1.825723 3.071879
H 2.011088 -3.000228 3.770618
N 1.492739 -1.982446 -4.771019
N 1.564693 -0.675518 -2.643655
C 1.556729 -1.275836 -3.654587
H 1.088381 -1.555568 -5.603815
H 2.168492 -2.730073 -4.905076
```

Ethylene glycol in water

55

C	0.55600	0.51500	-0.00000
C	-0.55600	-0.51500	-0.00000
O	-1.79500	0.19700	-0.00000
H	-2.51600	-0.44400	-0.00000
O	1.79500	-0.19700	0.00000
H	2.51600	0.44400	0.00000
H	-0.46600	-1.15100	-0.88800
H	-0.46600	-1.15100	0.88800
H	0.46600	1.15100	-0.88800
H	0.46600	1.15100	0.88800
O	1.91600	3.49200	1.62700
H	2.43000	4.20000	1.23800
H	1.58400	3.85800	2.44700
O	0.77700	-1.28000	4.16400
H	1.55800	-1.09900	3.64000
H	0.11300	-0.68500	3.81700
O	-5.43800	1.13500	-0.75200
H	-5.44300	1.26400	0.19600
H	-4.62300	0.66400	-0.92600
O	-1.66600	2.86700	1.14500
H	-0.97400	3.52400	1.06600
H	-1.49500	2.25100	0.43300
O	-2.81800	0.62900	-2.32100
H	-2.33200	0.77400	-1.50900
H	-3.22400	1.47400	-2.51100
O	-0.23200	-0.28200	-3.53400
H	-0.74700	0.49700	-3.74300
H	-0.28700	-0.82400	-4.32100
O	2.79100	-1.09000	2.51200
H	2.55700	-0.99600	1.58900
H	3.36400	-0.34500	2.69200
O	3.15100	1.76300	-1.48700
H	3.95200	1.28000	-1.28000
H	3.45700	2.55900	-1.92100
O	-1.93700	-2.41900	-2.88900
H	-1.00300	-2.60000	-2.78600
H	-2.01600	-1.47500	-2.75100
O	1.71800	-2.76500	-0.85700
H	1.32600	-2.87900	-1.72200
H	1.56700	-1.84300	-0.64500
O	4.10000	1.34600	1.89600
H	3.67100	2.15600	2.17200
H	4.84200	1.63600	1.36500
O	5.23900	0.15100	-1.01600
H	5.99900	0.44100	-1.51900
H	5.59100	-0.48700	-0.39500
O	-3.80600	-1.38400	0.33100
H	-4.17400	-0.92200	1.08400
H	-4.56400	-1.60400	-0.20900
O	1.03300	3.42700	-2.65800
H	1.87200	3.78100	-2.95200
H	1.26600	2.66100	-2.13300
O	-0.78500	1.10100	3.85500
H	-1.45000	1.17300	3.17000
H	-0.56200	2.00700	4.06900

NH₃ crystal

148

N	-0.181721	-0.181727	-0.181722
H	0.674578	0.141730	-0.634585
H	0.141729	-0.634584	0.674576
H	-0.634586	0.674581	0.141730
N	0.274291	-2.129822	2.225040
H	0.727154	-2.986121	2.548497
H	-0.582007	-2.453272	1.772184
H	-0.049161	-1.676957	3.081348
N	-2.129818	2.225046	0.274291
H	-2.453275	1.772183	-0.582008
H	-1.676962	3.081344	-0.049159
H	-2.986126	2.548498	0.727155
N	-2.200809	-2.654171	0.203298
H	-2.524266	-3.107033	-0.653001
H	-1.747952	-1.797872	-0.120152
H	-3.057117	-2.330718	0.656163
N	-2.654171	0.203304	-2.200813
H	-1.797872	-0.120153	-1.747951
H	-2.330722	0.656161	-3.057112
H	-3.107036	-0.653004	-2.524266
N	2.225048	0.274295	-2.129822
H	3.081347	-0.049162	-1.676959
H	2.548498	0.727152	-2.986121
H	1.772183	-0.582013	-2.453274
N	0.203300	-2.200815	-2.654177
H	0.656162	-3.057114	-2.330720
H	-0.652999	-2.524264	-3.107034
H	-0.120153	-1.747950	-1.797869
N	-2.583180	0.274297	2.678404
H	-1.726881	-0.049160	3.131267
H	-2.259730	0.727153	1.822105
H	-3.036045	-0.582011	2.354951
N	2.296039	0.345288	2.749396
H	3.152338	0.021831	3.202258
H	2.619489	0.798144	1.893097
H	1.843175	-0.511020	2.425943
N	0.345282	2.749394	2.296033
H	0.798145	1.893096	2.619490
H	-0.511016	2.425945	1.843177
H	0.021830	3.202259	3.152341
N	2.749401	2.296037	0.345282
H	2.425944	1.843174	-0.511017
H	3.202258	3.152335	0.021832
H	1.893093	2.619489	0.798147
N	2.678410	-2.583180	0.274289
H	2.354953	-3.036042	-0.582010
H	3.131267	-1.726881	-0.049160
H	1.822102	-2.259727	0.727154
N	0.274291	2.678402	-2.583184
H	0.727153	1.822103	-2.259728
H	-0.582008	2.354952	-3.036041
H	-0.049162	3.131266	-1.726876
N	-0.110729	-0.110734	4.697495
H	0.745569	0.212722	4.244633
H	0.212720	-0.563591	5.553794
H	-0.563594	0.745574	5.020948
N	-5.060940	-0.252718	-0.252714
H	-4.204641	0.070739	-0.705576
H	-4.737490	-0.705575	0.603585
H	-5.513805	0.603590	0.070739
N	-0.110730	4.697490	-0.110729

NH₃ crystal (cont.)

H	0.745569	5.020946	-0.563592	N	-2.058827	2.296038	5.153508
H	0.212720	4.244633	0.745569	H	-2.382283	1.843176	4.297209
H	-0.563595	5.553797	0.212723	H	-1.605970	3.152337	4.830058
N	4.697498	-0.110736	-0.110731	H	-2.915135	2.619491	5.606373
H	5.553797	0.212720	-0.563593	N	-2.129818	-2.583178	5.082515
H	5.020948	-0.563593	0.745568	H	-2.453274	-3.036040	4.226216
H	4.244633	0.745572	0.212722	H	-1.676961	-1.726879	4.759065
N	-0.252712	-5.060943	-0.252715	H	-2.986126	-2.259725	5.535380
H	0.603587	-4.737487	-0.705578	N	5.153510	-2.058831	2.296032
H	0.070738	-5.513800	0.603584	H	5.606373	-2.915130	2.619488
H	-0.705577	-4.204635	0.070738	H	4.297212	-2.382281	1.843175
N	-0.252712	-0.252720	-5.060940	H	4.830058	-1.605966	3.152340
H	0.603587	0.070737	-5.513802	N	-2.583180	5.082520	-2.129820
H	0.070737	-0.705576	-4.204641	H	-1.726881	4.759064	-1.676958
H	-0.705577	0.603588	-4.737487	H	-2.259731	5.535377	-2.986119
N	-4.604928	-2.200813	2.154049	H	-3.036045	4.226212	-2.453273
H	-4.152065	-3.057112	2.477505	N	2.296039	5.153511	-2.058829
H	-5.461226	-2.524262	1.701192	H	3.152338	4.830055	-1.605966
H	-4.928380	-1.747948	3.010357	H	2.619488	5.606368	-2.915128
N	2.154057	-4.604922	-2.200815	H	1.843174	4.297203	-2.382282
H	3.010356	-4.928378	-1.747952	N	5.082519	-2.129824	-2.583186
H	2.477507	-4.152065	-3.057113	H	5.535381	-2.986123	-2.259729
H	1.701192	-5.461230	-2.524267	H	4.226220	-2.453273	-3.036042
N	-2.725162	-4.675913	-2.271806	H	4.759066	-1.676959	-1.726878
H	-1.868863	-4.999369	-1.818944				
H	-2.401712	-4.223056	-3.128105				
H	-3.178027	-5.532220	-2.595259				
N	-4.675919	-2.271806	-2.725168				
H	-4.223057	-3.128105	-2.401712				
H	-5.532218	-2.595255	-3.178025				
H	-4.999372	-1.818941	-1.868860				
N	-2.200810	2.154053	-4.604927				
H	-2.524266	1.701190	-5.461226				
H	-1.747953	3.010351	-4.928376				
H	-3.057117	2.477505	-4.152062				
N	-2.271801	-2.725164	-4.675920				
H	-2.595257	-3.178026	-5.532219				
H	-1.818944	-1.868865	-4.999369				
H	-3.128108	-2.401711	-4.223055				
N	2.225048	-4.533929	2.678403				
H	3.081347	-4.857385	3.131265				
H	2.548498	-4.081072	1.822104				
H	1.772184	-5.390237	2.354950				
N	-2.654171	-4.604920	2.607411				
H	-1.797872	-4.928376	3.060274				
H	-2.330721	-4.152063	1.751113				
H	-3.107035	-5.461228	2.283959				
N	-4.533937	2.678403	2.225042				
H	-4.081074	1.822105	2.548498				
H	-5.390235	2.354954	1.772185				
H	-4.857389	3.131268	3.081350				
N	-4.604928	2.607411	-2.654175				
H	-4.152066	1.751112	-2.330719				
H	-5.461227	2.283961	-3.107032				
H	-4.928381	3.060275	-1.797867				
N	2.678409	2.225044	-4.533935				
H	2.354953	1.772181	-5.390234				
H	3.131266	3.081342	-4.857385				
H	1.822102	2.548496	-4.081071				
N	2.607419	-2.654173	-4.604928				
H	2.283962	-3.107035	-5.461227				
H	3.060275	-1.797874	-4.928378				
H	1.751111	-2.330720	-4.152063				

Formamide crystal

96

N	1.286359	2.999875	-1.104094	H	-0.688358	2.121114	-4.567767
C	0.242742	3.097622	-0.275868	H	-1.647082	1.125439	-1.898007
O	-0.802537	3.741221	-0.477515	H	0.409815	7.013402	-0.551300
H	2.097372	2.417793	-0.802447	H	1.158357	5.922672	0.658153
H	1.349389	3.508908	-2.011166	H	2.117533	6.918947	-2.012267
H	0.390106	2.512846	0.657860	H	4.249575	4.977292	-2.115380
N	-0.815909	5.044510	-2.806180	H	5.015934	4.540142	-3.675628
O	1.273682	4.302464	-3.432920	H	4.610358	2.986580	-1.133957
C	0.227952	4.945464	-3.633907	H	4.460181	-1.375156	-0.794546
H	-1.626678	5.625292	-3.107328	H	3.693262	-0.938392	0.764969
H	-0.878695	4.534177	-1.898608	H	4.099850	0.616155	-1.776629
H	0.081039	5.530838	-4.568295	H	-1.741377	4.454888	0.761706
O	-1.188677	1.022327	1.723869	H	-2.507600	4.890951	2.321060
C	-2.234408	1.665327	1.522882	H	-2.101464	6.444899	-0.219877
N	-3.279280	1.763388	2.350536	H	-4.089598	2.344769	2.048728
N	-0.750134	1.611767	-3.660122	H	-3.341055	1.254040	3.258181
C	-1.794446	1.710214	-2.831735	H	-2.382332	2.249716	0.588421
O	-2.839725	2.353813	-3.033383	H	1.786654	1.696769	3.040675
N	-2.852853	3.655802	-5.361547	H	2.553573	1.260005	1.481160
C	-1.809236	3.558055	-6.189774	H	2.146985	-0.294542	4.022759
O	-0.763505	2.915056	-5.988787	H	2.211827	3.589498	-4.671981
N	-4.249659	3.778488	-1.195625	H	2.978746	3.152734	-6.231496
C	-4.494403	4.980140	-1.725514	H	2.572610	1.598786	-3.690558
O	-5.040465	5.941515	-1.155370	H	-3.778429	3.066392	-1.795055
O	3.473971	0.714162	-5.310272	H	-4.544788	3.503543	-0.234807
C	2.928360	1.676136	-4.740789	H	-4.138653	5.057491	-2.775744
N	2.683616	2.877789	-5.270678	H	-3.664426	4.237498	-5.663928
O	-2.726891	6.543505	-3.754180	H	-2.915883	3.146769	-4.454475
C	-3.272954	7.504880	-3.184036	H	-1.956148	4.143430	-7.124162
N	-3.517833	8.707620	-3.713032	H	4.560744	5.698915	-5.959164
O	3.198037	1.500180	-0.156255	H	3.812202	6.789644	-7.168616
C	3.743648	0.538205	-0.725738	H	2.853027	5.793369	-4.498196
N	3.988844	-0.662848	-0.196509	H	-3.988927	9.418628	-3.114495
N	-2.212471	5.165896	1.360243	H	-3.222568	8.981478	-4.674744
C	-2.457215	6.367548	0.830354	H	-3.628144	7.427916	-2.133073
O	-3.003972	7.329624	1.400658	H	-1.315509	6.346916	-6.951111
N	-1.787298	7.058625	-6.352414	H	-2.081868	6.784066	-5.390863
C	-2.032042	8.260277	-6.882303	H	-1.676292	8.337628	-7.932533
O	-2.577653	9.222251	-6.312820				
N	2.257749	0.985761	2.442139				
C	2.503187	-0.216592	2.971867				
O	3.048798	-1.178566	2.402384				
O	5.511159	2.101570	-2.754404				
C	4.965548	3.063544	-2.184921				
N	4.720669	4.266284	-2.713916				
O	3.311430	5.690258	-0.876320				
C	2.264445	6.333572	-1.077879				
N	1.221279	6.431918	-0.250313				
O	1.233956	5.129330	2.078513				
C	2.280381	4.485629	2.279339				
N	3.324106	4.387670	1.452506				
O	1.659824	7.021357	-5.634305				
C	2.706114	6.378744	-5.432585				
N	3.749732	6.280998	-6.260811				
H	4.134424	3.806288	1.754315				
H	3.386577	4.896316	0.544701				
H	2.427853	3.900641	3.214461				
H	0.060184	1.030385	-3.358314				

Butanone in toluene

223

C	0.06600	0.88800	-0.37900	C	0.40100	4.18600	-1.68700
C	-0.86900	1.66300	0.52200	C	1.79600	4.29600	-1.63800
C	1.48800	0.70700	0.11900	C	2.55900	4.04000	-2.78400
H	-0.47300	2.66700	0.69900	H	0.04100	3.27800	-4.95600
H	-1.85600	1.73200	0.06900	H	-1.31600	3.73300	-2.92000
H	-0.94300	1.17400	1.49700	H	-0.19300	4.38500	-0.79600
C	2.38200	-0.08100	-0.83000	H	2.28700	4.58100	-0.70900
H	3.38800	-0.17800	-0.41900	H	3.64400	4.12600	-2.74600
H	1.97900	-1.08000	-1.00200	C	2.71800	3.40800	-5.16400
H	2.44900	0.41400	-1.80000	H	2.05700	3.14000	-5.98800
H	1.89700	1.70500	0.31300	H	3.28500	4.29900	-5.42900
H	1.43200	0.22700	1.10300	H	3.40400	2.58600	-4.96700
O	-0.30400	0.43700	-1.44200	C	3.35600	3.41500	2.47300
C	-1.81600	-2.25800	2.45700	C	3.91100	3.11500	3.72400
C	-2.82100	-1.31800	2.71600	C	3.08100	2.97800	4.84300
C	-3.71200	-0.94900	1.70100	C	1.69700	3.14000	4.71200
C	-3.59900	-1.52000	0.42800	C	1.14200	3.44000	3.46200
C	-2.59500	-2.46000	0.16900	C	1.97200	3.57700	2.34300
C	-1.70300	-2.83000	1.18400	H	4.98800	2.98900	3.82500
H	-2.90800	-0.87400	3.70600	H	3.51200	2.74500	5.81500
H	-4.49300	-0.21700	1.90200	H	1.05100	3.03400	5.58200
H	-4.29200	-1.23300	-0.36100	H	0.06500	3.56600	3.36000
H	-2.50700	-2.90500	-0.82100	H	1.54000	3.81000	1.37000
H	-0.92200	-3.56100	0.98200	C	4.21600	3.55700	1.31400
C	-0.89300	-2.64100	3.50800	H	5.25200	3.39700	1.60900
H	-1.13500	-2.09700	4.41900	H	3.93500	2.82100	0.56200
H	-0.97500	-3.71100	3.69100	H	4.10600	4.55900	0.90200
H	0.12400	-2.40200	3.20100	C	-4.84800	1.11100	4.67200
C	-7.75700	-0.97300	1.70200	C	-5.34500	1.85200	3.59300
C	-7.35200	0.16300	0.99000	C	-4.55400	2.84400	3.00100
C	-8.14300	1.31700	1.00500	C	-3.26500	3.09400	3.48700
C	-9.33900	1.33700	1.73300	C	-2.76800	2.35300	4.56600
C	-9.74400	0.20100	2.44400	C	-3.55900	1.36100	5.15800
C	-8.95300	-0.95300	2.42900	H	-6.34800	1.65700	3.21500
H	-6.42200	0.14700	0.42400	H	-4.94100	3.42000	2.16200
H	-7.82800	2.20100	0.45100	H	-2.65000	3.86500	3.02700
H	-9.95500	2.23500	1.74500	H	-1.76500	2.54800	4.94400
H	-10.67400	0.21700	3.01000	H	-3.17200	0.78500	5.99700
H	-9.26800	-1.83700	2.98300	C	-5.66700	0.08400	5.28500
C	-6.93700	-2.16900	1.68500	H	-6.62900	0.03500	4.77600
H	-6.05500	-1.99600	1.07200	H	-5.16400	-0.87800	5.20400
H	-6.63100	-2.41300	2.70200	H	-5.82500	0.32400	6.33500
H	-7.51100	-2.99700	1.27100	C	5.46800	-4.12400	-4.31300
C	-6.82100	-2.16000	-2.40700	C	4.18400	-4.22000	-4.86200
C	-5.90200	-1.70400	-3.36000	C	3.07200	-3.81400	-4.11500
C	-6.08600	-0.45800	-3.97000	C	3.24300	-3.31300	-2.81900
C	-7.18900	0.33300	-3.62800	C	4.52800	-3.21700	-2.27000
C	-8.10800	-0.12200	-2.67500	C	5.64000	-3.62300	-3.01700
C	-7.92400	-1.36900	-2.06500	H	4.05000	-4.61000	-5.87000
H	-5.04400	-2.32000	-3.62700	H	2.07300	-3.88900	-4.54200
H	-5.37100	-0.10400	-4.71200	H	2.37800	-2.99800	-2.23800
H	-7.33200	1.30300	-4.10200	H	4.66200	-2.82800	-1.26200
H	-8.96600	0.49300	-2.40800	H	6.63900	-3.54900	-2.59000
H	-8.63900	-1.72300	-1.32400	C	6.62100	-4.54400	-5.08700
C	-6.63000	-3.45100	-1.77500	H	6.29200	-4.90200	-6.06100
H	-5.72600	-3.91600	-2.16600	H	7.29600	-3.70000	-5.21800
H	-6.53400	-3.31700	-0.69900	H	7.13800	-5.34500	-4.56100
H	-7.48700	-4.08900	-1.98700	C	1.70300	-1.80900	6.97400
C	1.92700	3.67300	-3.97800	C	1.54900	-1.26900	8.25700
C	0.53200	3.56300	-4.02700	C	0.94100	-0.01900	8.42200
C	-0.23100	3.81900	-2.88200	C	0.48700	0.69100	7.30400

Butanone in toluene (cont.)

C	0.64100	0.15200	6.02200	H	-3.69100	0.63100	9.75100
C	1.24900	-1.09800	5.85700	H	-6.06300	-0.05700	10.05800
H	1.90300	-1.82200	9.12600	H	-7.00600	-1.91600	8.69900
H	0.82200	0.40100	9.42000	H	-5.57700	-3.08800	7.03200
H	0.01400	1.66400	7.43300	C	-2.99700	-2.57100	6.48300
H	0.28800	0.70400	5.15300	H	-1.99900	-2.13600	6.49100
H	1.36900	-1.51800	4.85900	H	-2.93400	-3.62800	6.73900
C	2.33300	-3.10400	6.80300	H	-3.43300	-2.46300	5.49100
H	2.61800	-3.50000	7.77600	C	-6.51400	4.16200	-1.41100
H	1.63400	-3.78700	6.32300	C	-5.69100	5.22200	-1.01300
H	3.22000	-2.99700	6.18100	C	-4.31300	5.02200	-0.86600
C	2.31500	-3.93400	1.98000	C	-3.75800	3.76100	-1.11600
C	2.20300	-4.91700	2.97100	C	-4.58100	2.70100	-1.51400
C	2.00900	-6.25600	2.61000	C	-5.95900	2.90100	-1.66100
C	1.92600	-6.61200	1.25800	H	-6.12200	6.20300	-0.81800
C	2.03800	-5.62900	0.26800	H	-3.67300	5.84700	-0.55600
C	2.23200	-4.29000	0.62800	H	-2.68600	3.60500	-1.00200
H	2.26800	-4.64100	4.02200	H	-4.14900	1.72000	-1.70900
H	1.92200	-7.02100	3.38000	H	-6.59900	2.07600	-1.97100
H	1.77500	-7.65300	0.97800	C	-7.94100	4.36900	-1.56300
H	1.97300	-5.90500	-0.78400	H	-8.18400	5.40400	-1.32600
H	2.31900	-3.52500	-0.14200	H	-8.23100	4.15300	-2.59000
C	2.51600	-2.54800	2.35400	H	-8.47800	3.70600	-0.88700
H	2.55400	-2.46600	3.43900	C	-2.03100	-7.78600	-1.38600
H	3.45300	-2.19000	1.93000	C	-2.48500	-7.63400	-2.70200
H	1.69200	-1.94600	1.97300	C	-3.62600	-8.32200	-3.13200
C	-5.71800	-5.76400	1.03100	C	-4.31200	-9.16200	-2.24700
C	-6.33300	-5.44200	2.24700	C	-3.85800	-9.31400	-0.93200
C	-7.53700	-6.05800	2.60700	C	-2.71700	-8.62600	-0.50100
C	-8.12700	-6.99600	1.75200	H	-1.95100	-6.98000	-3.39000
C	-7.51300	-7.31800	0.53500	H	-3.97900	-8.20400	-4.15500
C	-6.30800	-6.70200	0.17500	H	-5.20000	-9.69700	-2.58200
H	-5.87300	-4.71300	2.91300	H	-4.39200	-9.96700	-0.24300
H	-8.01500	-5.80800	3.55300	H	-2.36400	-8.74400	0.52200
H	-9.06400	-7.47500	2.03200	C	-0.84900	-7.07300	-0.94000
H	-7.97200	-8.04700	-0.13000	H	-0.45500	-6.47500	-1.76100
H	-5.83100	-6.95200	-0.77100	H	-1.11100	-6.42100	-0.10800
C	-4.47100	-5.12600	0.65800	H	-0.09400	-7.78800	-0.61800
H	-4.16900	-4.43300	1.44100				
H	-3.70000	-5.88500	0.52700				
H	-4.60600	-4.58200	-0.27600				
C	-1.40300	-1.02300	-4.87200				
C	-0.64300	-1.79800	-3.98900				
C	-0.71400	-3.19500	-4.04400				
C	-1.54600	-3.81700	-4.98300				
C	-2.30600	-3.04200	-5.86600				
C	-2.23500	-1.64500	-5.81100				
H	0.00500	-1.31400	-3.25900				
H	-0.12200	-3.79800	-3.35700				
H	-1.60100	-4.90400	-5.02600				
H	-2.95400	-3.52600	-6.59600				
H	-2.82700	-1.04200	-6.49800				
C	-1.32900	0.42400	-4.81500				
H	-0.63700	0.72100	-4.02900				
H	-0.97900	0.80800	-5.77200				
H	-2.31700	0.83000	-4.60100				
C	-3.83000	-1.88800	7.45400				
C	-3.29900	-0.84200	8.21900				
C	-4.10300	-0.18300	9.15600				
C	-5.43800	-0.56900	9.32900				
C	-5.96800	-1.61500	8.56400				
C	-5.16400	-2.27500	7.62700				
H	-2.26100	-0.54200	8.08500				

A.3 CspLOX2 Model System Structures

Given is the planar structure of the 2,5-heptadienyl model with oxygen approaching to C13. The atoms of the substrate (small model) in the large model structure of the CspLOX2 wild type active site are shown in red, the additional ones from the iron moiety (medium model) in blue.

2,5-heptadienyl + O₂

20

C	16.978013	7.777701	24.683949
C	16.479215	9.176118	24.508856
C	15.137170	9.409339	24.328861
C	14.225898	8.324404	24.309190
C	12.841853	8.564924	24.123562
C	11.958234	7.512912	24.104488
C	12.457031	6.114495	24.279580
H	17.698113	7.514463	23.888218
H	17.486095	7.660451	25.658181
H	16.171582	7.023040	24.649595
H	13.552280	6.059435	24.415317
H	11.999372	5.641962	25.167435
H	12.211389	5.495973	23.397472
H	10.888700	7.698776	23.961043
H	17.183410	10.014512	24.524057
H	12.476513	9.589182	23.995318
H	14.771830	10.433598	24.200617
H	14.590571	7.302016	24.437201
O	16.313260	9.290388	25.894281
O	17.617686	9.093809	26.565271

large CspLOX2 model

294

C	24.315	54.378	64.407
C	24.113	53.191	63.459
C	25.249	53.021	62.375
C	25.047	51.940	61.283
C	23.892	52.185	60.402
C	23.921	51.477	59.087
C	24.413	50.300	58.631
C	24.690	49.057	59.345
C	24.616	47.848	58.547
C	24.986	46.540	58.801
C	25.491	45.880	60.066
C	25.301	46.809	61.263
C	25.687	46.174	62.602
C	27.252	46.248	62.744
C	27.743	45.951	64.151
C	27.733	44.451	64.535
C	28.649	44.186	65.715
C	28.182	42.898	66.415
O	28.323	41.816	65.810
O	27.702	42.994	67.552
H	23.914	52.265	63.999
H	23.157	53.481	63.023
H	25.357	54.004	61.915
H	26.206	52.728	62.808
H	25.924	52.020	60.640
H	24.947	51.056	61.913
H	22.995	51.915	60.959
H	23.875	53.263	60.241
H	23.349	51.920	58.285
H	24.385	50.157	57.551
H	24.206	48.946	60.304
H	24.102	48.010	57.600
H	24.345	45.943	58.169
H	25.060	44.915	60.331
H	26.542	45.711	59.833
H	24.259	47.120	61.335
H	25.847	47.744	61.137
H	25.419	45.118	62.630
H	25.231	46.842	63.333
H	27.617	47.247	62.502
H	27.641	45.515	62.037
H	27.390	46.697	64.864
H	28.806	46.190	64.151
H	28.023	43.847	63.676
H	26.761	44.067	64.845
H	28.730	45.035	66.394
H	29.617	43.956	65.270
H	25.301	54.427	64.868
H	23.516	54.539	65.130
H	24.280	55.311	63.844

large model (cont.)

Fe	27.361	50.230	55.309	O	20.430	47.960	50.704
O	26.960	49.884	57.219	C	22.044	49.434	53.413
H	27.523	49.207	57.594	C	23.475	49.261	53.770
H	26.190	49.896	57.788	C	24.385	50.227	54.042
C	29.018	47.800	55.914	N	24.065	48.017	54.083
O	28.471	48.627	55.153	C	25.320	48.267	54.426
O	28.706	47.799	57.099	N	25.521	49.606	54.490
N	29.621	46.232	54.059	H	21.522	51.558	51.842
C	29.889	46.573	55.432	H	22.681	48.961	51.384
H	29.678	47.017	53.426	H	21.453	50.188	53.934
H	28.691	45.843	54.002	H	21.522	48.572	53.828
C	31.434	46.852	55.539	H	23.629	47.121	53.918
H	29.560	45.726	56.034	H	24.192	51.290	54.019
H	31.923	45.883	55.432	H	26.045	47.528	54.734
H	31.476	48.617	54.294	H	19.586	49.270	52.032
H	33.032	47.809	54.361	H	21.703	50.862	50.366
H	31.848	47.078	53.419	N	22.540	56.472	54.628
C	31.956	47.640	54.347	C	23.236	55.197	54.683
H	31.602	48.376	57.031	C	22.219	53.986	54.584
H	31.423	46.744	57.801	O	22.115	53.398	53.541
C	33.399	47.278	57.219	C	24.127	55.079	55.918
H	33.674	46.233	57.074	C	25.132	53.937	55.963
H	33.810	47.825	56.370	C	25.686	53.184	54.979
H	33.713	47.615	58.207	N	25.839	53.720	57.093
C	31.871	47.320	56.992	C	26.745	52.718	56.924
O	29.186	51.118	55.860	N	26.524	52.236	55.663
C	30.278	51.690	55.609	H	22.677	57.114	55.395
N	31.377	51.236	56.061	H	23.899	55.167	53.818
H	31.476	50.304	56.438	H	24.781	55.951	55.902
H	32.260	51.704	55.915	H	23.491	55.149	56.800
C	30.314	52.910	54.688	H	25.671	54.137	57.998
H	29.330	52.922	54.219	H	25.355	53.293	53.957
H	31.088	52.703	53.950	H	27.481	52.366	57.632
C	30.505	54.238	55.490	N	21.521	53.652	55.632
H	29.944	54.274	56.424	H	21.573	54.273	56.427
C	31.949	54.283	55.994	C	20.677	52.432	55.781
O	32.187	53.824	57.124	H	21.373	51.602	55.657
H	32.745	54.705	55.380	C	19.487	52.424	54.811
N	30.140	55.392	54.644	O	19.348	51.396	54.108
H	30.708	55.495	53.815	H	18.800	53.267	54.738
H	30.231	56.232	55.197	C	20.252	52.327	57.198
N	27.944	54.313	48.713	H	21.137	52.569	57.786
C	28.144	53.563	49.984	H	19.454	53.049	57.370
C	29.457	54.013	50.716	C	19.740	50.986	57.604
O	29.477	54.541	51.827	H	18.930	50.762	56.910
C	28.021	52.086	49.823	C	19.151	50.937	59.089
C	28.293	51.294	51.045	H	18.732	49.988	59.425
C	27.627	51.379	52.217	H	19.988	51.074	59.774
N	29.318	50.394	51.262	H	18.578	51.864	59.118
C	29.361	50.081	52.588	C	20.723	49.810	57.495
N	28.313	50.615	53.216	H	20.657	49.338	56.515
H	29.956	49.977	50.600	H	21.744	50.158	57.647
H	26.833	52.068	52.462	H	20.699	49.007	58.231
H	30.015	49.336	53.016	H	21.551	56.275	54.572
H	26.975	51.874	49.601	N	22.926	58.172	59.203
H	28.583	51.703	48.971	C	22.783	56.764	59.385
H	27.389	53.839	50.720	C	21.417	56.049	59.112
H	30.408	53.848	50.210	O	21.283	54.878	58.821
H	28.136	53.848	47.837	C	23.345	56.453	60.791
H	28.538	55.129	48.753	C	24.829	56.812	61.002
N	22.032	50.858	51.321	C	25.919	55.979	60.326
C	21.862	49.481	51.881	O	27.137	56.139	60.675
C	20.488	48.893	51.550	O	25.598	55.242	59.359

large model (cont.)

H	22.934	58.717	60.054	C	30.582	50.935	66.143
H	23.529	56.228	58.798	C	31.204	51.727	65.001
H	22.748	56.892	61.590	O	31.513	51.141	63.958
H	23.254	55.369	60.867	C	29.120	51.207	66.365
H	25.031	57.870	60.840	C	28.114	51.085	65.152
H	25.050	56.648	62.057	C	28.154	49.651	64.598
H	20.512	56.652	59.190	C	26.653	51.300	65.658
H	22.157	58.494	58.632	H	30.912	51.525	68.234
N	25.081	41.405	54.848	H	30.715	49.896	65.840
C	25.207	42.629	55.588	H	28.839	50.498	67.144
C	23.881	43.159	56.058	H	29.025	52.214	66.773
O	23.764	43.640	57.209	H	28.296	51.936	64.495
C	26.166	43.608	54.924	H	27.522	49.500	63.723
C	26.587	44.789	55.814	H	29.185	49.554	64.258
C	25.674	44.171	53.578	H	27.788	48.918	65.317
C	27.373	44.355	56.964	H	26.457	50.749	66.578
H	25.240	41.564	53.864	H	26.584	52.370	65.856
H	25.696	42.435	56.543	H	26.002	51.091	64.809
H	27.028	42.989	54.676	H	31.375	52.799	65.102
H	27.130	45.507	55.200	H	32.198	51.341	67.275
H	25.686	45.341	56.081	N	34.336	49.892	63.909
H	26.459	44.771	53.117	C	33.857	48.666	63.245
H	25.387	43.329	52.948	C	33.439	48.834	61.773
H	24.803	44.796	53.772	O	33.765	48.035	60.916
H	28.115	43.565	56.845	C	32.843	47.898	64.133
H	27.854	45.248	57.364	C	32.217	46.643	63.489
H	26.671	44.018	57.727	O	33.342	47.615	65.419
H	23.017	43.125	55.394	H	33.976	50.160	64.814
H	24.151	41.033	54.978	H	34.692	47.966	63.259
N	22.095	43.688	59.778	H	31.996	48.574	64.258
C	21.177	44.761	60.160	H	33.541	48.447	65.854
C	19.849	44.201	60.754	H	31.829	46.918	62.508
O	19.413	44.757	61.767	H	32.963	45.849	63.492
C	20.932	45.657	58.885	H	31.305	46.439	64.051
H	22.440	43.621	58.831	H	34.139	50.658	63.281
H	21.661	45.317	60.963	N	32.613	49.842	61.545
H	20.407	46.561	59.193	C	32.192	50.369	60.292
H	21.915	45.891	58.477	C	33.360	50.704	59.348
H	20.356	45.138	58.119	O	33.305	50.302	58.145
H	19.330	43.359	60.297	C	31.216	51.527	60.435
H	21.622	42.818	59.977	C	29.804	51.140	60.856
N	21.075	45.286	64.531	C	28.939	52.294	61.251
C	20.924	46.772	64.427	C	28.995	50.413	59.757
C	19.504	47.189	64.822	H	32.379	50.329	62.398
O	19.315	48.069	65.667	H	31.780	49.540	59.716
C	21.299	47.376	63.119	H	31.510	52.132	61.293
C	20.963	48.973	63.065	H	31.220	52.009	59.458
C	21.888	49.790	63.960	H	29.847	50.417	61.670
C	21.292	49.539	61.638	H	28.146	51.833	61.841
H	20.740	44.772	63.728	H	29.409	52.971	61.964
H	21.698	47.227	65.044	H	28.427	52.865	60.477
H	22.349	47.138	62.952	H	29.020	50.973	58.823
H	20.751	46.984	62.262	H	29.492	49.452	59.624
H	19.901	49.084	63.285	H	27.993	50.293	60.168
H	21.708	50.862	63.882	H	34.217	51.275	59.705
H	21.516	49.474	64.935				
H	22.921	49.484	63.793				
H	22.336	49.323	61.411				
H	20.585	49.110	60.929				
H	20.984	50.584	61.676				
H	18.650	46.703	64.351				
H	20.562	44.973	65.343				
N	31.270	50.988	67.457				

A.4 Code Segment of the Steric Shielding Analysis

A small segment of the Perl script carrying out the sphere model for steric shielding analysis is provided. Given that the x-, y- and z-coordinates of all atoms are stored in the correct order in arrays of the form @coordx, @coor dy and @coor dz, the segment computes the accessibility *A* on the example of the C11 carbon center (\$C11acc) for one MD snapshot.

```
# calculation of the three points P1, P2 and P3
# for plane construction by centers of mass
$P1x=($coordx[$C9]+$coor dx[$C10]+$coor dx[$C11])/3;
$P1y=($coor dy[$C9]+$coor dy[$C10]+$coor dy[$C11])/3;
$P1z=($coor dz[$C9]+$coor dz[$C10]+$coor dz[$C11])/3;
$P2x=($coor dx[$C10]+$coor dx[$C11]+$coor dx[$C12])/3;
$P2y=($coor dy[$C10]+$coor dy[$C11]+$coor dy[$C12])/3;
$P2z=($coor dz[$C10]+$coor dz[$C11]+$coor dz[$C12])/3;
$P3x=($coor dx[$C11]+$coor dx[$C12]+$coor dx[$C13])/3;
$P3y=($coor dy[$C11]+$coor dy[$C12]+$coor dy[$C13])/3;
$P3z=($coor dz[$C11]+$coor dz[$C12]+$coor dz[$C13])/3;

# normal vector of the plane by vector product
$normx=(( $P3y-$P1y)*($P2z-$P1z)-($P3z-$P1z)*($P2y-$P1y));
$normy=(( $P3z-$P1z)*($P2x-$P1x)-($P3x-$P1x)*($P2z-$P1z));
$normz=(( $P3x-$P1x)*($P2y-$P1y)-($P3y-$P1y)*($P2x-$P1x));
$norm=sqrt($normx**2+$normy**2+$normz**2);

# sanity check to ensure that the vector points to the substrate
# side opposed to the iron center by dot product
$ncos=($normx*($coor dx[$C11]-$coor dx[$Fe])+$normy*($coor dy[$C11],
-$coor dy[$Fe])+$normz*($coor dz[$C11]-$coor dz[$Fe]));
if ($ncos < 0) {
    $normx=-$normx;
    $normy=-$normy;
    $normz=-$normz;
}

# find the points along the normal with 3 Angstroms distance to C11
$PC11x=$coor dx[$C11]+3*$normx/$norm;
$PC11y=$coor dy[$C11]+3*$normy/$norm;
$PC11z=$coor dz[$C11]+3*$normz/$norm;

# large starting value for minimal distance
$PC11_mindist=1000;

# loop over all atoms to find the distance to the closest one
foreach $i (1..$n) {
    $PC11_dist=sqrt(($coor dx[$i]-$PC11x)**2+($coor dy[$i]-$PC11y)**2,
+($coor dz[$i]-$PC11z)**2);
    if ($PC11_dist < $PC11_mindist) {
        $PC11_mindist=$PC11_dist;
    }
}

# compare minimal distance to rmax and rmin
if ($PC11_mindist >= $rmax) {
    $C11acc=1;
}
elseif (($PC11_mindist < $rmax) && ($PC11_mindist > $rmin)) {
    $C11acc=(1/($rmax-$rmin))*($PC11_mindist-$rmin),
*exp($PC11_mindist-$rmax);
}
else {
    $C11acc=0;
}
}
```


Bibliography

- [1] J. Kaminsky, R. A. Mata, H.-J. Werner and F. Jensen, *Mol. Phys.* **2008**, *106*, 1899.
- [2] H.-J. Werner and M. Schütz, *J. Chem. Phys.* **2011**, *135*, 144116.
- [3] C. Riplinger and F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106.
- [4] T. Yoshikawa and H. Nakai, *A Linear-Scaling Divide-and-Conquer Quantum Chemical Method for Open-Shell Systems and Excited States*, in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems* (ed M. S. Gordon), chapter 10, John Wiley & Sons, Ltd, Chichester, UK, **2017**.
- [5] H.-J. Werner, C. Köppl, Q. Ma and M. Schwilk, *Explicitly Correlated Local Electron Correlation Methods*, in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems* (ed M. S. Gordon), chapter 1, John Wiley & Sons, Ltd, Chichester, UK, **2017**.
- [6] F. Pavošević, C. Peng, P. Pinski, C. Riplinger, F. Neese and E. F. Valeev, *J. Chem. Phys.* **2017**, *146*, 174108.
- [7] M. S. Gordon, D. G. Federov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.* **2012**, *112*, 632.
- [8] M. A. Collins and R. P. A. Bettens, *Chem. Rev.* **2015**, *115*, 5607.
- [9] C. Köppl and H.-J. Werner, *J. Chem. Theory Comput.* **2016**, *12*, 3122.
- [10] T. Shimizu, *Ann. Rev. Pharmacol. Toxicol.* **2009**, *49*, 123–150.
- [11] A. Andreou and I. Feussner, *Phytochemistry* **2009**, *70*, 1504.
- [12] M. Born and R. Oppenheimer, *Ann. d. Phys.* **1927**, *389*, 457.
- [13] I. N. Levine, *Quantum Chemistry*, chapter 13.1, Pearson Prentice Hall, sixth edition, **2009**.
- [14] J. C. Slater, *Phys. Rev.* **1929**, *34*, 1293.
- [15] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, chapter 1.3.1, Dover Publications, **1996**.
- [16] H.-J. Werner, *Advanced Methods in Quantum Chemistry*, University Lecture, **2014**, University of Stuttgart.
- [17] Z. Tóth and P. Pulay, *J. Chem Phys.* **2016**, *145*, 164102.

- [18] J. M. Foster and S. F. Boys, *Rev. Mod. Phys.* **1960**, *32*, 300–302.
- [19] J. Pipek and P. G. Mezey, *J. Comput. Phys.* **1989**, *90*, 4916.
- [20] G. Knizia, *J. Chem. Theory Comput.* **2013**, *9*, 4834.
- [21] P. Hohenberg and W. Kohn, *Phys. Rev. B* **1964**, *136*, 864.
- [22] S. F. Sousa, P. A. Fernandes and M. J. Ramos, *J. Phys. Chem. A* **2007**, *111*, 10439–10452.
- [23] A. J. Cohen, P. Mori-Sánchez and W. Yang, *Science* **2008**, *321*, 792.
- [24] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- [25] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
- [26] J. C. Slater, *Phys. Rev.* **1951**, *81*, 385.
- [27] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
- [28] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
- [29] S. H. Vosko, L. Wilk and M. Nusair, *J. Phys.* **1980**, *58*, 1200.
- [30] S. Grimme, J. Antony, S. Ehrlich and S. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [31] S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456.
- [32] E. Schwegler, M. Challacombe and M. Head-Gordon, *J. Chem. Phys.* **1997**, *106*, 9708.
- [33] C. Ochsenfeld, C. A. White and M. Head-Gordon, *J. Chem. Phys.* **1998**, *109*, 1663.
- [34] F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.* **2009**, *356*, 98.
- [35] S. Kossmann and F. Neese, *Chem. Phys. Lett.* **2009**, *481*, 240.
- [36] R. Izsak and F. Neese, *J. Chem. Phys.* **2011**, *135*, 144105.
- [37] R. Polly, H.-J. Werner, F. R. Manby and P. J. Knowles, *Mol. Phys.* **2004**, *102*, 2311.
- [38] R. A. Kendall and H. A. Früchtl, *Theor. Chem. Acc.* **1997**, *97*, 158.
- [39] S. Hamel, M. E. Casida and D. R. Salahub, *J. Chem. Phys.* **2001**, *114*, 7342.

- [40] F. Weigend, *Phys. Chem. Chem. Phys.* **2002**, 4, 4285.
- [41] A. Sodt and M. Head-Gordon, *J. Chem. Phys.* **2008**, 128, 104106.
- [42] F. Aquilante, T. B. Pedersen and R. Lindh, *J. Chem. Phys.* **2007**, 126, 194106.
- [43] B. I. Dunlap, J. W. D. Connolly and J. R. Sabin, *J. Chem. Phys.* **1979**, 71, 3396.
- [44] F. R. Manby, *J. Chem. Phys.* **2003**, 119, 4607.
- [45] M. Schütz and F. R. Manby, *Phys. Chem. Chem. Phys.* **2003**, 5, 3349.
- [46] S. S. Xantheas, *J. Chem. Phys.* **1996**, 104, 8821.
- [47] K. Mierzwicki and Z. Latajka, *Chem. Phys. Lett.* **2003**, 380, 654–664.
- [48] C. D. Sherrill, *Counterpoise correction and basis set superposition error*, **2010**, <http://vergil.chemistry.gatech.edu/notes/cp.pdf>.
- [49] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt and J. H. van Lenthe, *Chem. Rev.* **1994**, 94, 1873.
- [50] D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.* **2003**, 119, 3599.
- [51] D. W. Zhang, Y. Xiang, A. M. Gao and J. Z. H. Zhang, *J. Chem. Phys.* **2004**, 120, 1145.
- [52] Y. Xiang, D. W. Zhang and J. Z. H. Zhang, *J. Phys. Chem. B* **2003**, 107, 12039.
- [53] X. He and J. Z. H. Zhang, *J. Chem. Phys.* **2006**, 124, 184703.
- [54] X. Wang, J. Liu, J. Z. H. Zhang and X. He, *J. Phys. Chem. A* **2013**, 117, 7149.
- [55] J. Liu, T. Zhu, X. He and J. Z. H. Zhang, *MFCC-Based Fragmentation Methods for Biomolecules*, in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems* (ed M. S. Gordon), chapter 11, John Wiley & Sons, Ltd, Chichester, UK, **2017**.
- [56] W. Guo, A. Wu and X. Xu, *Chem. Phys. Lett.* **2010**, 498, 203.
- [57] K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, *Chem. Phys. Lett.* **1999**, 313, 701.
- [58] M. S. Gordon, J. M. Mullin, S. R. Pruitt, L. B. Roskop, L. V. Slipchenko and J. A. Boatz, *J. Phys. Chem. B* **2009**, 113, 9646.

- [59] D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A* **2012**, *116*, 704.
- [60] D. G. Fedorov and K. Kitaura, *Modeling and Visualization for the Fragment Molecular Orbital Method with the Graphical User Interface FU, and Analyses of Protein–Ligand Binding*, in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems* (ed M. S. Gordon), chapter 3, John Wiley & Sons, Ltd, Chichester, UK, **2017**.
- [61] D. G. Fedorov and K. Kitaura, *J. Chem. Phys.* **2009**, *131*, 171106.
- [62] P. A. Kollman, *AMBER 11*, University of California San Francisco, **2011**.
- [63] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [64] G. S. Grest and K. Kremer, *Phys. Rev. A* **1986**, *33*, 3628.
- [65] R. A. Mata and H. Stoll, *Chem. Phys. Lett.* **2008**, *465*, 136.
- [66] M. Werner and R. A. Mata, *AIP Conf. Proc.* **2015**, *1702*, 090069.
- [67] T. H. Dunning, Jr., *J. Chem. Phys.* **1989**, *90*, 1007.
- [68] F. Weigend, *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.
- [69] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O’Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson and M. Wang, *MOLPRO, version 2012.1, a package of ab initio programs*, **2012**, see <http://www.molpro.net>.
- [70] W. J. Hehre, R. F. Stewart and J. A. Pople, *J. Chem. Phys.* **1969**, *51*, 2657.
- [71] J. B. Collins, P. v. R. Schleyer, J. S. Binkley and J. A. Pople, *J. Chem. Phys.* **1976**, *64*, 5142.
- [72] P. A. Bash, M. J. Field and M. Karplus, *J. Am. Chem. Soc.* **1987**, *109*, 8092.

- [73] Z. Yang and Y.-P. Zhao, *Materials Science and Engineering A* **2006**, *423*, 84.
- [74] H. M. Senn and W. Thiel, *Angew. Chem.* **2009**, *121*, 1220.
- [75] L. Tian and R. A. Friesner, *J. Chem. Theory Comput.* **2009**, *5*, 1421.
- [76] A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425.
- [77] R. Bjornsson and M. Bühl, *J. Chem. Theory Comput.* **2012**, *8*, 498.
- [78] Y. Mao, Y. Shao, J. Dziedzic, C.-R. Skylaris, T. Head-Gordon and M. Head-Gordon, *J. Chem. Theory Comput.* **2017**, *13*, 1963–1979.
- [79] A. Ganguly, E. Boulanger and W. Thiel, *J. Chem. Theory Comput.* **2017**, *13*, 2954–2961.
- [80] P. König, M. Hoffmann and Th. Frauenheim, *J. Phys. Chem.* **2005**, *109*, 9082.
- [81] K. Senthilkumar, J. I. Mujika, K. E. Ranaghan, F. R. Manby, A. J. Mulholland and J. N. Harvey, *J. R. Soc. Interface* **2008**, *5*, 207.
- [82] G. G. Ferenczy, *J. Comput. Chem.* **2013**, *34*, 854.
- [83] G. G. Ferenczy, *J. Comput. Chem.* **2013**, *34*, 862.
- [84] P. Cortona, *Phys. Rev. B* **1991**, *44*, 8454.
- [85] T. A. Wesolowski and A. Warshel, *J. Phys. Chem.* **1993**, *97*, 8050.
- [86] J. M. H. Olsen, N. H. List, K. Kristensen and J. Kongsted, *J. Chem. Theory Comput.* **2015**, *11*, 1832.
- [87] J. M. H. Olsen, C. Steinmann, K. Ruud and J. Kongsted, *J. Phys. Chem. A* **2015**, *119*, 5344.
- [88] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [89] C. Møller and M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.
- [90] C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269.
- [91] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
- [92] A. E. Reed, R. B. Weinstock and F. Weinhold, *J. Comput. Phys.* **1985**, *83*, 735.

- [93] W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657.
- [94] S. Grimme, *J. Chem. Phys.* **2003**, *118*, 9095.
- [95] A. Wuttke and R. A. Mata, *J. Comput. Chem.* **2017**, *38*, 15.
- [96] S. Saebø and P. Pulay, *Chem. Phys. Lett.* **1985**, *113*, 13.
- [97] P. Pulay and S. Saebø, *Theor. Chim. Acta* **1986**, *69*, 357.
- [98] C. Hampel and H.-J. Werner, *J. Chem. Phys.* **1996**, *104*, 6286.
- [99] M. Schütz, *J. Chem. Phys.* **2002**, *116*, 8772.
- [100] R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz and Y. Cao, *J. Phys. Chem. A* **1999**, *103*, 1913.
- [101] M. S. Lee, P. E. Maslen and M. Head-Gordon, *J. Chem. Phys.* **2000**, *112*, 3592.
- [102] D. Mejía-Rodríguez and A. M. Köster, *J. Chem. Phys.* **2014**, *141*, 124114.
- [103] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györfy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson and M. Wang, *MOLPRO, version 2015.1, a package of ab initio programs*, **2015**, see <http://www.molpro.net>.
- [104] R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724.
- [105] A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.* **1994**, *100*, 5829.
- [106] J. Chieffari, Y. K. Chong, F. Ercole, J. Krstina, J. Jeffery, T. P. T. Le, R. T. A. Mayadunne, G. F. Meijs, C. L. Moad, G. Moad, E. Rizzardo and S. H. Thang, *Macromolecules* **1998**, *31*, 5559–5562.

- [107] W. Meiser, *Investigation of the Kinetics and Mechanism of RAFT Polymerization via EPR Spectroscopy*, PhD thesis, University of Göttingen, **2012**, <http://d-nb.info/1044074043/34>.
- [108] H. T. T. Duong, N. N. M. Adnan, N. Barraud, J. S. Basuki, S. K. Kutty, K. Jung, N. Kumar, T. P. Davis and C. Boyer, *J. Mater. Chem. B* **2014**, 2, 5003.
- [109] M. Buback, T. Tunkers and P. Vana, *Macromol. Rapid Commun.* **2005**, 26, 796.
- [110] M. Muntaz, K. Aissou, D. Katsigiannopoulos, C. Brochon, E. Cloutet and G. Hadziioannou, *RSC Adv.* **2015**, 5, 98559.
- [111] A. Klamt and G. Schürmann, *J. Chem. Soc. Perkin Trans.* **1993**, 5, 799.
- [112] H.-J. Werner, F. R. Manby and P. J. Knowles, *J. Chem. Phys.* **2003**, 118, 8149.
- [113] P. Söderhjelm and U. Ryde, *J. Phys. Chem. A* **2009**, 113, 617.
- [114] P. Söderhjelm, F. Aquilante and U. Ryde, *J. Phys. Chem. B* **2009**, 113, 11085.
- [115] A. Andreou, F. Brodhun and I. Feussner, *Prog. Lipid Res.* **2009**, 48, 148.
- [116] T. Horn, S. Adel, R. Schumann, S. Sur, K. R. Kakularam, A. Polamarasetty, P. Redanna, H. Huhn and D. Heydeck, *Prog. Lipid Res.* **2015**, 57, 13.
- [117] J. S. Wright, H. Shadnia and L. L. Chepelev, *Biochemistry* **2009**, 30, 1016.
- [118] M. J. Knapp, K. Rickert and J. P. Klinman, *J. Am. Chem. Soc.* **2002**, 124, 3865.
- [119] M. H. M. Olsson, P. E. M. Siegbahn and A. Warshel, *J. Am. Chem. Soc.* **2004**, 126, 2820.
- [120] E. Hatcher, A. V. Soudackov and S. Hammes-Schiffer, *J. Am. Chem. Soc.* **2004**, 126, 5763.
- [121] I. Tejero, L. A. Eriksson, À. González-Lafont, J. Marquet and J. M. Lluch, *J. Phys. Chem. B* **2004**, 108, 13831.
- [122] E. Hatcher, A. V. Soudackov and S. Hammes-Schiffer, *J. Am. Chem. Soc.* **2007**, 129, 187.

- [123] M. Barroso, L. G. Arnaut and S. J. Formosinho, *J. Phys. Org. Chem.* **2008**, *21*, 659.
- [124] S. J. Edwards, A. V. Soudackov and S. Hammes-Schiffer, *J. Phys. Chem. B* **2010**, *114*, 6653.
- [125] A. V. Soudackov and S. Hammes-Schiffer, *J. Phys. Chem. Lett.* **2014**, *5*, 3274.
- [126] M. Hamberg, C. Su and E. Oliw, *J. Biol. Chem.* **1998**, *273*, 13080–13088.
- [127] A. Wennman, A. Magnuson, M. Hamberg and E. H. Oliw, *J. Lipid. Res.* **2015**, *56*, 1606.
- [128] A. Andreou, C. Göbel, M. Hamberg and I. Feussner, *J. Biol. Chem.* **2010**, *285*, 14178.
- [129] T. Borowski and E. Broclawik, *J. Phys. Chem. B* **2003**, *107*, 4639.
- [130] C. Schneider, D. A. Pratt, N. A. Porter and A. R. Brash, *Chemistry & Biology* **2007**, *14*, 473.
- [131] M. E. Newcomer and A. R. Brash, *Prot. Sci.* **2015**, *24*, 298.
- [132] J. Newie, P. Neumann, M. Werner, R. A. Mata, R. Ficner and I. Feussner, *Sci. Rep.* **2017**, *7*, 2069.
- [133] D. A. Pratt, J. H. Mills and N. A. Porter, *J. Am. Chem. Soc.* **2003**, *125*, 5801.
- [134] D. Hu and D. A. Pratt, *Chem. Commun.* **2010**, *46*, 3711.
- [135] M. J. Knapp, F. P. Seebeck and J. P. Klinman, *J. Am. Chem. Soc.* **2001**, *123*, 2931.
- [136] K. E. Furse, D. A. Pratt, C. Schneider, A. R. Brash, N. A. Porter and T. P. Lybrand, *Biochemistry* **2006**, *45*, 3206.
- [137] F. Neese, *WIREs Comput. Mol. Sci.* **2012**, *2*, 73.
- [138] P. Li, B. P. Roberts, D. K. Chakravorty and K. M. Merz, *J. Chem. Theory Comput.* **2013**, *9*, 2733.
- [139] U. C. Singh and P. A. Kollman, *J. Comp. Chem.* **1984**, *5*, 129.
- [140] A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.* **1992**, *97*, 2571.
- [141] K. A. Tallman, D. A. Pratt and N. A. Porter, *J. Am. Chem. Soc.* **2001**, *123*, 11827.

-
- [142] E. H. Oliw, *Prostaglandins Other Lipid Mediat.* **2002**, 68-69, 313.
- [143] F. Libisch, C. Huang and E. A. Carter, *Acc. Chem. Res.* **2014**, 47, 2768.
- [144] E. A. Ivanova-Shor, A. M. Shor, S. S. Laletina, R. Ch. Deka and V. A. Naslutov, *Russ. Chem. Bull.* **2017**, 66, 759.