New Exploratory and Spectroscopic Tools for Computational Transition Metal Chemistry

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

FLORIAN STEFAN KRAUSBECK

MSc Interdisciplinary Sciences, ETH Zurich

born 13.02.1987

citizen of Germany

accepted on the recommendation of

Prof. Dr. Markus Reiher, examiner
Prof. Dr. Sereina Riniker, co-examiner

2017
Dedicated to my parents and my brother.
## Contents

**Abstract**  
ix  
**Zusammenfassung**  
xi  
1 Introduction  
1  
2 Theory  
5  
2.1 Foundations of Quantum Chemistry  
5  
2.1.1 The Schrödinger Equation  
5  
2.1.2 The Born Oppenheimer Approximation  
7  
2.1.3 The Harmonic Approximation  
8  
2.1.4 The Slater Determinant Approximation  
9  
2.1.5 The Hartree–Fock Approximation  
10  
2.1.6 The Roothaan–Hall Equations  
11  
2.2 Density Functional Theory  
12  
2.2.1 The Hohenberg–Kohn Theorems  
13  
2.2.2 Kohn–Sham Density Functional Theory  
14  
2.3 Gradient-driven Molecule Construction  
16  
2.4 Frozen-Density Embedding  
20  
2.5 Vibrational Raman Optical Activity Spectroscopy  
21  
3 Molecular Scaffold Design by GdMC  
27  
3.1 Generating a Molecular Environment by Frozen-Density Embedding  
31  
3.1.1 Challenges with the FDE Starting Guess  
32  
3.2 Constructing the Chemical Environment by Atomwise Additions  
35  
3.3 The Design Target: CO₂ Activation  
36  
3.4 Protocol for Shell-wise Construction of an Embedding Environment  
36  
3.4.1 Binding Sites and Coordination Geometry  
37  
3.4.2 Scoring and Constrained Optimization  
38  
3.4.3 Multidentate Ligands and Ring Structures  
39  
3.4.4 Structure Tree and Structural Inheritance  
40  
3.4.5 Pruning of the Structure Tree  
42
Contents

3.4.6 Graphical Explanation of the MSD .......................... 43
3.5 Constructing the Target Complex ............................... 45
  3.5.1 Position of the Target Branch in the Screened Molecular Space .......................... 49
  3.5.2 Functionalization as a Means to Further Lower the Fragment Gradient .................. 51
3.6 Designing Alternative Candidate Structures .................. 51
3.7 Comparison with Similar Ligand Environments ................. 55
4 Conditional Design of Molecular Scaffolds ....................... 59
  4.1 Principles of Conditional Design ............................... 61
    4.1.1 The Generic Network ..................................... 63
    4.1.2 Shell-wise Construction of Molecular Scaffolds .......... 64
    4.1.3 Structure Trees and the Split and Prune Method ........ 65
    4.1.4 The Property Classes ................................... 66
    4.1.5 Conditional Design for a Generic Chemical-Process Network .......................... 66
    4.1.6 Expression of Conditions in Terms of Local Descriptors .................. 68
  4.2 Practical Realization of Conditional Design .................. 69
  4.3 Generic Network for Dinitrogen Fixation ..................... 70
  4.4 Conditional Design of Schrock-type Scaffolds ................. 72
    4.4.1 Preparations for an Extensive Screening of the Schrock-Scaffold Space .............. 73
    4.4.2 Extensive Screening of the Schrock-Scaffold Space .................................. 75
    4.4.3 Continuation of the Design Process at the NH$_3$ Node .............................. 80
    4.4.4 The Löwdin Partial Charge as Valid Conditional Descriptor ................ 81
5 Stereoisomer Discrimination by RVROA .......................... 83
  5.1 VROA and RVROA of H$_2$O$_2$ ................................ 86
  5.2 VROA and RVROA of (S)-methylxirane ......................... 89
  5.3 VROA and RVROA Spectra of Naproxen and Ibuprofen ........ 93
  5.4 VROA and RVROA for Micro-Solvated Carboxylate Anions ..... 101
    5.4.1 RVROA for the Naproxen Carboxylate Anion .......................... 102
    5.4.2 VROA/RVROA for the Ibuprofen Carboxylate Anion .......................... 105
Contents

6 First Application of RVROA to Chiral Metal Complexes  109
   6.1 Preparatory Steps Towards RVROA Spectra  . . . . . . . . . .  111
   6.2 RVROA Spectra for the Neutral Fe-Helicene Complex .  115
   6.3 Towards RVROA Spectra for the Oxidized Fe-Helicene
       Complex  . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .  118
7 Conclusions and Perspectives  121
   A Computational Methodology  127
   B Publications  135
   Acknowledgments  137

Bibliography  137
Abstract

A reactive fragment which is structurally unstable can be stabilized through binding to a suitable transition-metal fragment. The metal fragment may be designed in a shell-wise build-up of a surrounding molecular environment. However, adding more and more atoms in an iterative fashion soon leads to a combinatorial explosion, which is unfeasible to handle without automation. This thesis is devoted to the conceptual and algorithmic development of a fully automated and parallelized framework called "molecular scaffold designer" (MSD) that constructs the embedding environment atom-wise. Molecular realizations of such an environment are formed based on heuristic rules and subsequently optimized by electronic structure methods. Structures are evaluated within the concept of gradient-driven molecule construction that introduces structure modifications to reduce the forces on all atoms. For all generated structures, the nuclear gradient on the reactive fragment and its coordination energy are evaluated to steer the design process. We develop and analyze our approach at the example of CO₂ activation by reproducing a known compound and mapping out possible alternative structures and their effect on the stabilization of the CO₂ ligand.

Standard quantum chemical calculations can reliably predict molecular properties for a given structural configuration. By contrast, molecular design approaches search for structures that feature pre-defined properties. Hence, the desired direction from property to structure contrasts the viable direction from structure to property. Consequently, massive computational screening for structures with sought-for properties has become a convenient solution owing to ever cheaper and increasingly powerful computer hardware. However, as a combinatorial problem, exhaustive searches are still unfeasible as the chemical compound space is vast. At the same time, molecular structures usually have to fulfill more than one constraint in a given context. For example, in a catalytic cycle energetic and structural-stability constraints on various intermediates
ABSTRACT

need to be fulfilled by the same molecular scaffold. Based on the MSD, we explore options to constrain structural design of molecules with pre-defined function under multiple constraints. We denote this approach conditional design and apply it to the construction of chelate-ligand scaffolds in transition-metal mediated catalysis.

Determining the absolute configuration of a molecule is possible by means of various chiroptical spectroscopic methods, one of them being vibrational Raman optical activity (VROA). By adopting a laser excitation wavelength in resonance with an electronic transition, the weak spectral signals can be selectively enhanced. In the second part of this thesis, we implement a Kohn–Sham methodology for the calculation of resonance and off-resonance VROA spectra. With this new implementation, we calculate the resonance and off-resonance VROA spectra for hydrogen peroxide, methyloxirane, (S)-naproxen, and (S)-ibuprofen and discuss their band patterns. The resonance-enhancement of the VROA spectrum of (S)-naproxen at an incident wavelength of 514.5 nm caused by an absorption tail as well as the typical off-resonance behavior of the VROA spectrum of (S)-ibuprofen at the same incident wavelength can be well reproduced. VROA spectra are also predicted under full resonance conditions.

After having tested our implementation at the example of those smaller molecules, we then investigate the resonance VROA spectra for an ethynyl-helicene mono-Fe^{II} complex grafted with one electroactive [Fe(κ^2-dppe) (η^3-C_5Me_5)] fragment which was found to exhibit unprecedented redox-triggered chiroptical switching. Based on our calculations, we predict a strong dependence of the RVROA band of the alkynyl stretching mode on the incident wavelength λ. The experimental RVROA spectrum was well reproduced with respect to the relative RVROA intensities and the normal modes responsible for the corresponding RVROA signature could be readily identified. Furthermore, we carefully discuss and analyze what challenges have to be faced when calculating RVROA spectra for such transition metal complexes.
Zusammenfassung


Gewöhnliche quantenchemische Berechnungen sind in der Lage, molekulare Eigenschaften verlässlich und genau vorherzusagen. Im Gegen- satz dazu suchen Ansätze für Molekül-design nach Strukturen, welche eine vorgegebene Eigenschaft erfüllen. Dies bedeutet jedoch, dass die erwünschte Vorgehensweise ‚von Eigenschaft auf Struktur‘ dem bisherigen Ansatz ‚von Struktur auf Eigenschaft‘ gegenübersteht. Im Zuge dessen wurden aufwändige und rechenintensive Studien in die Wege ge-
ZUSAMMENFASSUNG


Introduction

Understanding the influence of atomic compositions of molecular structures on chemical reactivity is at the heart of quantum chemistry [1–3]. As a direct consequence, it has firmly established itself in many areas of chemical research, not only as a complement to experimental studies, but also as a method to predict experimental outcomes, e.g., a specific reaction pathway possibly not considered before [4–7] or certain band patterns of a molecular spectrum [8].

To this end, a variety of different quantum chemical methods has been developed over the last few decades [9, 10], such as highly accurate and computationally demanding approaches based on a wave function ansatz, e.g., coupled cluster [11–14], many-body perturbation theory [15], multireference methods such as configuration interaction [16], perturbation approaches [17], or the recently developed density matrix renormalization group method [18–24].

Clearly, it is desirable to apply the most accurate method whenever possible, for example, to study molecular systems containing one or more transition metal atoms [25, 26]. While there are ongoing efforts to make larger and more elaborate systems accessible by these approaches [27, 28], density functional theory (DFT) [29–32] with its favorable linear scaling [33–35] is still the method of choice for the study of transition metal complexes [36]. Prominent examples for computational investigations with DFT of transition metal systems and how certain structural features change a complex’ property include the work by Ziegler and co-workers [37–41], the group of Frenking [42–46], or the studies by Siegbahn and co-workers [47–50].
In general, it is a formidable task to design transition metal complexes exhibiting a certain physical property. Work along those lines has led to new concepts and ideas for rational compound design [51], e.g., Quantitative Structure–Activity Relationships or Quantitative Structure–Property Relationships [52–57]. However, from a conceptual point of view, such high-throughput screening approaches [58] could in principle be rendered obsolete by going the opposite way, that is, from property to structure, which involves inverting the Schrödinger equation. Already in 1929, Abrahazumian studied how a differential equation is defined by its eigenvalue spectrum [59]. It was soon proven that the inversion is generally not possible [60–62] since similar molecular properties can be described by completely different Hamiltonians, i.e., completely different molecular structures. However, several solution strategies for this ill-posed problem [63] have emerged, such as inverse perturbation analysis [64–69], the optimized wave function approach [70, 71] or the linear combination of atom-centered potentials [72–75]. In chapter 3, we describe a fully automated approach to generate new molecular structures for a transition metal complex based on a new design principle, namely the concept of gradient-driven molecule construction (GdMC) [76, 77] which has recently been introduced by our group. The theoretical foundations needed to understand this design principle are introduced in detail in chapter 2. We apply this approach to the design of a CO₂ activating catalyst.

In the general case, a molecular scaffold has to fulfill specific and different roles at different nodes of a network of molecular-rearrangement steps, i.e., a catalytic cycle. To illustrate this in more detail, we consider the case of dinitrogen fixation, for which a catalytic cycle at a single metal center has been formulated by Chatt in 1976 [78, 79]. In 2003, Yandulov and Schrock were then able to present the first catalyst that can accomplish the task under ambient conditions [80, 81]. The Schrock catalyst is able to coordinate dinitrogen sufficiently strongly, while the coordination of ammonia is not too strong so that dinitrogen can eventually replace ammonia. In chapter 4, we introduce the concept of conditional design at the example of the Chatt-Schrock cycle [82, 83], i.e., we specifically address the problem that optimizing a molecular scaffold to fulfill properties of a single node in a network cannot be carried out independently of optimizing this type of molecular scaffold for all other nodes in the same network.
Besides designing a molecular scaffold which exhibits a certain predefined property and which is able to facilitate several molecular functions in a catalytic cycle, it is essential to identify such a structure in an experimental setting as well. To this end, several spectroscopic methods have been developed, one of them being vibrational Raman optical activity (VROA) [84, 85]. Besides other chiroptical techniques, e.g., electronic and vibrational circular dichroism and optical rotation [86, 87], RVROA has become the predominant technique to determine the absolute configuration of a molecule [84, 88–90] and to identify conformational changes of chiral biomolecules [91, 92]. Here, theoretical studies are essential to assign an observed spectral feature to a specific molecular vibration and structure [93–97]. Due to the very weak VROA intensities and other potential side effects such as fluorescence [98, 99], it is desirable to selectively enhance VROA signals by adopting a laser excitation wavelength in resonance with an electronic transition. In chapter 5, we implement a new approach for calculating such resonance VROA spectra (VROA) and benchmark our implementation at the example of several smaller molecules. In chapter 6, we then apply the same methodology to the calculation of RVROA spectra for an ethynyl-helicene mono-Fe\textsuperscript{II} complex and address potential challenges which can arise in the theoretical RVROA study of transition metal complexes. Finally, we conclude this work in chapter 7 and present an outlook.
In this chapter, we give a short overview of the quantum mechanical concepts and equations employed in this work. After reviewing the basics of quantum chemistry and the Hartree–Fock approximation, we introduce the equations governing density functional theory. We then introduce the working equations for gradient-driven molecule construction and outline possible solution strategies, e.g., involving frozen density embedding techniques. A summary of the theory for off-resonance and resonance vibrational Raman Optical Activity spectroscopy closes this chapter.

Throughout this thesis, we present scalar and unspecified quantities in normal type, vectors and matrices in boldface. Also, we will in general adopt the same symbol for classical quantities and their corresponding quantum mechanical operators. All equations are expressed in Hartree atomic units, that is, $1/(4\pi\varepsilon_0) = \hbar = m_e = 1$.

2.1 Foundations of Quantum Chemistry

2.1.1 The Schrödinger Equation

In non-relativistic quantum mechanics [100–102], the state of a quantum system and its change in time is governed by the time-dependent Schrödinger equation [103]

$$\hat{H}\Psi(r, t) = i\hbar \frac{\partial}{\partial t} \Psi(r, t). \quad (2.1)$$
It is a partial differential equation with the Hamilton operator $\hat{H}$ and the wave function $\Psi$ which fully describes the state of the quantum system at time $t$. The mathematical expression for the Hamilton operator $\hat{H}$ as well as any other quantum mechanical operator can be derived by means of the correspondence principle [102].

For the studies to follow, the so called time-independent Schrödinger equation is of primary importance which can be obtained by a separation of variables in Eq. (2.1), if the Hamilton operator $\hat{H}$ does not explicitly depend on time $t$:

$$\hat{H}\Psi(r) = E\Psi(r) \tag{2.2}$$

The Hamiltonian $\hat{H}$ as the total energy operator can be expressed as a sum of a kinetic operator $\hat{T}$ and a potential energy operator $\hat{V}$,

$$\hat{H} = \hat{T} + \hat{V}, \tag{2.3}$$

where $\hat{T}$ has the following form:

$$\hat{T} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{M} \frac{1}{2M_i} \nabla_i^2. \tag{2.4}$$

Here, $M$ is the number of nuclei and $N$ the number of electrons. $M_i$ represents the mass of nucleus $I$ and $\nabla_i$ or $\nabla_I$ the gradient with respect to the coordinates of particle $i$ or $I$, respectively.

The potential energy operator $\hat{V}$ is given by

$$\hat{V} = -\sum_{i=1}^{N} \sum_{j=1}^{M} \frac{Z_i}{|r_i - R_j|} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} + \sum_{i=1}^{M} \sum_{j>i}^{M} \frac{Z_i Z_j}{R_{ij}}, \tag{2.5}$$

with $Z_i$ being the charge number of nucleus $I$, $R_I$ the position of nucleus $I$, $r_i$ the position of electron $i$, $R_{ij} = |R_i - R_j|$ the distance between nuclei $I$ and $J$, and $r_{ij} = |r_i - r_j|$ the distance between electrons $i$ and $j$. Thus, we treat all nuclei and all electrons as point-like particles and account for the attractive interaction between nuclei and electrons as well as the mutual repulsion between all nuclei and all electrons among themselves.

Solving the Schrödinger equation for molecular properties is the paramount task of any quantum chemical calculation. In the following, we introduce several approximations which facilitate the solution of this complicated equation.
2.1.2 The Born Oppenheimer Approximation

The highly complex form of the Hamiltonian \( \hat{H} \) which in turn leads to the very complicated Schrödinger equation requires approximate solutions to be able to treat not only single particles, but larger chemical systems of interest. One such approximation was introduced by Born and Oppenheimer in 1927 [104] who were the first to separate nuclear and electronic coordinates. This approach simplifies the Schrödinger equation and is justified as the nuclear masses are in general several orders of magnitude larger than the mass of an electron. In essence, the electrons move much faster than nuclei and are able to instantaneously adapt to the nuclei changing their position. We can thus define the electronic Hamiltonian by subtracting the kinetic energy of the nuclei from the full Hamiltonian:

\[
\hat{H}_{el} = \hat{H} - \sum_{i=1}^{M} \frac{1}{2M_i} \nabla_i^2. \tag{2.6}
\]

As a result, the nuclear coordinates can be set to fixed values and treated as parameters in the electronic Schrödinger equation. The latter can thus be solved for a specific nuclear configuration \( \{ R_i \} \) and the energy eigenvalues for all possible nuclear configurations make up the so-called potential energy surface (PES).

\[
\hat{H}_{el}(\{ r_i \}; \{ R_i \}) \Psi_{el}(\{ r_i \}; \{ R_i \}) = E(\{ R_i \}) \Psi_{el}(\{ r_i \}; \{ R_i \}). \tag{2.7}
\]

Here, the electronic energy \( E(\{ R_i \}) \) and the electronic wave function \( \Psi_{el}(\{ r_i \}; \{ R_i \}) \) now depend on a specific nuclear configuration \( \{ R_i \} \) introduced as parameters into the equation.

Subsequently, the total state function denoted by the index \( k \) can be expanded in the basis of electronic state functions denoted by the index \( n \), that is,

\[
\Psi_k(\{ r_i \}; \{ R_i \}) = \sum_{n} \chi_{k,n}(\{ R_i \}) \Psi_{el,n}(\{ r_i \}; \{ R_i \}). \tag{2.8}
\]

The complete Schrödinger equation then takes the following form:

\[
- \sum_{i} \frac{1}{2M_i} \nabla_i^2 \chi_{k,m} + \sum_{n} C_{m,n} \chi_{k,m} + E_{el,m} \chi_{k,m} = E_{k} \chi_{k,m}. \tag{2.9}
\]
with \( C_{m,n} \) being the so called nonadiabatic couplings [105] representing couplings between near-lying vibrational levels of different electronic states which are defined as

\[
C_{m,n} = \sum_I 2 \langle \Psi_{el,m} | \nabla_I | \Psi_{el,n} \rangle \left( -\frac{1}{2M_I} \nabla_I \right) + \langle \Psi_{el,m} | \left( -\frac{1}{2M_I} \nabla_I^2 \right) | \Psi_{el,n} \rangle.
\]

(2.10)

These couplings are set to zero in the Born–Oppenheimer approximation such that the nuclear coordinates adopt some fixed values. As a result, we obtain

\[
- \sum_I \frac{1}{2M_I} \nabla_I^2 \chi_{k,m} + E_{el,m} \chi_{k,m} \approx E_k \chi_{k,m}.
\]

(2.11)

### 2.1.3 The Harmonic Approximation

In the expression above, the first term represents the kinetic energy of all nuclei, while the second term represents the potential energy which is nothing else than the electronic energy. Within the Born–Oppenheimer approximation, the electronic energy can only be calculated for individual nuclear configurations. In practice, the electronic energy \( E_{el} \) is expanded into a Taylor series around the molecular equilibrium structure and truncated after the second order term. After subtracting the constant zero-order term and neglecting the linear term which is also zero for a molecular structure in its equilibrium, we obtain the so called harmonic approximation

\[
E_{el,m}^{(vib)} \approx \frac{1}{2} \sum_{I,J} \sum_{\alpha,\beta} r_{I,\alpha} \left( \frac{\partial^2 E_{el,m}}{\partial r_{I,\alpha} \partial r_{J,\beta}} \right) r_{J,\beta} \quad \alpha, \beta \in \{x, y, z\}.
\]

(2.12)

Writing Eq. (2.12) in matrix form and including the individual atomic masses gives us the so called mass-weighted Hessian matrix \( H^{(m)} \) with the elements

\[
H_{Ia,I\beta}^{(M)} = \frac{1}{\sqrt{M_I M_J}} \left( \frac{\partial^2 E_{el,m}}{\partial r_{I,\alpha} \partial r_{J,\beta}} \right).
\]

(2.13)

Diagonalizing the mass-weighted Hessian by means of a unitary transformation decouples the corresponding differential equations and yields
new and decoupled coordinates, known as normal modes describing the collective motion of all nuclei for a given vibration. The corresponding independent differential equations resemble the form of the harmonic oscillator which can be solved straightforwardly [106]. The resulting eigenvalues of the mass-weighted Hessian matrix are the squares of the harmonic angular frequencies.

### 2.1.4 The Slater Determinant Approximation

As a first expression for the electronic state function $\Psi_{\text{el}}$, we adopt a product of single particle functions or one electron wave functions which is also known as the Hartree product:

$$\Psi_{\text{el}}^{\text{HP}}(x_1, x_2, \ldots, x_N) = \chi_1(x_1) \chi_2(x_2) \ldots \chi_K(x_N) \quad (2.14)$$

The $\chi$'s are constructed to be orthonormal, i.e. $\langle \chi_i | \chi_j \rangle = \delta_{ij}$, and are a product of a spatial part $\psi$ and a spin part, i.e.,

$$\chi(x) = \psi(r) \cdot \begin{pmatrix} \alpha(\omega) \\ \beta(\omega) \end{pmatrix} \quad (2.15)$$

It should be noted that the electron spin $\alpha$ or $\beta$ has to be introduced artificially as it does not arise in a natural fashion in non-relativistic quantum chemistry [107]. Electrons are particles with half-integer spin, i.e., with a spin quantum number of $s = 1/2$, and are able to adopt two spin states, called $|\alpha\rangle = \alpha(\omega)$ for the state $M_s = +1/2$ and $|\beta\rangle = \beta(\omega)$ for the state $M_s = -1/2$, respectively.

We point out that the Hartree product is not an eigenfunction of a two-particle operator, that is, of the Coulomb term in the electronic Hamiltonian $\hat{H}_{\text{el}}$ and it does not respect the Pauli antisymmetry principle [94].

However, by rewriting the Hartree product as a matrix determinant, also known as Slater determinant [108], it respects this important constraint:

$$\Psi^{\text{SD}}(x_1, x_2, \ldots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \cdots & \chi_N(x_N) \end{vmatrix}. \quad (2.16)$$
2.1.5 The Hartree–Fock Approximation

Based on the Slater determinant, we can now derive the so-called Hartree–Fock equations \([109–112]\) which are a one-determinant ansatz to the exact wave function \(\Psi_\text{el}\).

For this, we first calculate the energy for an arbitrary trial wave function \(\tilde{\Psi}_\text{el}\) as the expectation value of the Hamiltonian:

\[
E[\tilde{\Psi}_\text{el}] = \langle \tilde{\Psi}_\text{el} | \hat{H} | \tilde{\Psi}_\text{el} \rangle. \tag{2.17}
\]

By minimizing the trial wave function according to the variational principle \([113, 114]\), the exact ground state energy \(E_0^{\text{exact}}\) can be approximated from above:

\[
E_0 = \min_{\delta E_0 \to 0} E[\tilde{\Psi}_\text{el}] \geq E_0^{\text{exact}}. \tag{2.18}
\]

One choice for the trial wave function \(\Phi\) is the Slater determinant \(\Psi^{SD}(x_1, x_2, \ldots, x_N)\). Taking into account the constraint of orthonormal orbitals, the Lagrange-multiplier method \([115]\) can be adopted to minimize the energy functional by means of varying the spin orbitals. Introducing the Lagrange multipliers as \(\epsilon_{ab}\) and the bracket notation \(\langle a | b \rangle\) for \(\int dx_1 \chi_a(x_1) \chi_b(x_1)\), we then have:

\[
\mathcal{L}[\{\chi_a\}] = E_0[\{\chi_a\}] - \sum_{a=1}^{N} \sum_{b=1}^{N} \epsilon_{ab} (\langle a | b \rangle - \delta_{ab}) \tag{2.19}
\]

Applying a unitary transformation yields the so-called canonical Hartree–Fock equations:

\[
\begin{align*}
&\left[ -\frac{1}{2} \nabla^2 \sum_{A} \frac{Z_A}{r_{1A}} \right] \chi_a(1) + \sum_{b} \left[ \int \! dx_2 \chi_b(2) \frac{1}{r_{12}} \chi_b(2) \right] \chi_a(1) \\
&\quad - \sum_{b} \left[ \int \! dx_2 \chi_b(2) \frac{1}{r_{12}} \chi_b(2) \right] \chi_b(1) = \epsilon_a \chi_a, \tag{2.20}
\end{align*}
\]

with \(\epsilon_a\) being the energy of spin orbital \(\chi_a\).

We now introduce several shorthand notations to simplify Eq. (2.20).

Abbreviating the coordinate of electron \(i\) with \(x_i = i\), we can write for the one-electron core Hamiltonian:
\( h(1) \chi_a + \sum_b J_b(1) \chi_a(1) - \sum_b K_b(1) \chi_a(1) = \epsilon_a \chi_a(1), \quad (2.21) \)

The one-electron Coulomb operator we abbreviate as:

\( J_b(1) \chi_a(1) = \left[ \int d\mathbf{x}_2 \chi_b(2) \frac{1}{r_{12}} \chi_a(2) \right] \chi_a(1) \quad (2.22) \)

and, finally, the one-electron exchange operator \( K_b(1) \) as:

\( K_b(1) \chi_a(1) = \left[ \int d\mathbf{x}_2 \chi_b(2) \frac{1}{r_{12}} \chi_a(2) \right] \chi_b(1). \quad (2.23) \)

Written together, those three operators make up the so called Fock operator:

\( f(1) = h(1) + \sum_b J_b - K_b \quad (2.24) \)

resulting in a simplified notation for the Hartree–Fock equations:

\( f(1) \chi_a = \epsilon_a \chi_a. \quad (2.25) \)

The Hartree–Fock equation can be solved self-consistently, as the Fock operator implicitly depends on all orbitals \( \{ \chi_a \} \).

### 2.1.6 The Roothaan–Hall Equations

For closed-shell molecules, i.e., only doubly occupied spin orbitals, we can integrate out the orthonormal spin functions as the following expression holds:

\( \langle \alpha | \beta \rangle = \delta_{\alpha \beta} \quad (2.26) \)

With this in mind, the Fock operator becomes:

\( f = h + \sum_{b=1}^{N/2} 2J_b - K_b \quad (2.27) \)

which is now applied on a set of spatial orbitals \( \{ \psi_i(\mathbf{r}) \} \):

\( f \psi_a = \epsilon_a \psi_a \quad (2.28) \)
In molecular calculations, the orbitals are expanded employing atom-centered basis functions,

\[ \psi_j = \sum_v^K C_{vj} \phi_v. \]  \hspace{6cm} (2.29)

We now insert Eq. (2.29) into Eq. (2.28) and multiply by \( \phi^*_\mu \) before integrating over the whole space which results in:

\[ \sum_v^K C_{vj} \int dr_1 \phi^*_\mu(1) f(1) \phi_v(1) = \epsilon_j \sum_v^K C_{vj} \int dr_1 \phi^*_\mu(1) \phi_v(1). \]  \hspace{6cm} (2.30)

We can now define the Fock Matrix \( F \) as:

\[ F_{\mu\nu} = \int dr_1 \phi^*_\mu(1) f(1) \phi_\nu(1), \]  \hspace{6cm} (2.31)

and the so called overlap matrix \( S \) as:

\[ S_{\mu\nu} = \int dr_1 \phi^*_\mu(1) \phi_\nu(1), \]  \hspace{6cm} (2.32)

and express Eq. (2.30) as an equation in matrix form which is known as the Roothaan–Hall equation [116, 117]:

\[ FC = SC\epsilon. \]  \hspace{6cm} (2.33)

### 2.2 Density Functional Theory

As an alternative to applying a wave function-based approach such as the Hartree–Fock method to calculate molecular properties, density functional theory (DFT) has received much attention over the last decades [29, 118, 119]. Here, the state function is discarded and instead focus is laid on the electron density itself as a function of space instead of all electron coordinates \( \{x_i\} \), i.e., a physical quantity which is computationally much more feasible. The goal is now to express the electronic energy by the electron density

\[ \rho(r) = N_{el} \int \cdots \int |\Psi_0(r_1, r_2, \ldots, r_{N_{el}})|^2 dr_2 \cdots dr_{N_{el}} \]  \hspace{6cm} (2.34)
Density Functional Theory | 2.2

with $N_{el}$ representing the total number of electrons in the studied system and the subscript "o" indicates that we are only interested in the ground-state electron density.

### 2.2.1 The Hohenberg–Kohn Theorems

Already in the early 1920’s, models have been constructed by Thomas and Fermi [120–122] relating the energy of a homogenous electron gas to its electronic density. Yet, only the theoretical works by Hohenberg and Kohn in 1964 made an application to more general chemical systems possible [73].

The first Hohenberg–Kohn theorem relates the ground-state density $\rho_0$ of a system to the ground-state wave function $\Psi_0$ by mapping the electron density to the external potential defining the system:

$$\rho \rightarrow \mathcal{V}_{\text{ext}}$$

(2.35)

As a result, the energy can be expressed as a functional of the electron density

$$E_v[\rho] = \langle \Psi[\rho] | \hat{H} | \Psi[\rho] \rangle = \int \rho(r) \, dr + F_{\text{HK}}[\rho]$$

(2.36)

with the so called Hohenberg–Kohn functional $F_{\text{HK}} = T[\rho] + V_{\text{ee}}[\rho]$ which is system-independent.

A strategy for solving this equation is given by the second Hohenberg–Kohn theorem, i.e.,

$$E_0 = \min_{\rho \in \{\rho\}} E_v[\rho].$$

(2.37)

In essence, the variational principle is applied in order to minimize the energy for the ground state $E_0$ over the set of all possible electron densities $\{\rho\}$. The fixed external potential of the system is indicated by $V_0$. 
### 2.2.2 Kohn–Sham Density Functional Theory

Applying the Euler–Lagrange formalism [123] to minimize the energy in a variational fashion subject to the constraint that the integral of the electron density equals the number of electrons yields

\[ \delta \left\{ E_v[\rho] - \mu \left[ \int \rho(r) \, dr - N \right] \right\} = 0 \]  \hspace{1cm} (2.38)

with

\[ \mu = \frac{\delta E_v[\rho]}{\delta \rho(r)} = v(r) + \frac{\delta F_{HK}[\rho]}{\delta \rho(r)}. \]  \hspace{1cm} (2.39)

Here, we lack an expression for the kinetic-energy functional \( T[\rho] \). In 1965, Kohn and Sham [124] considered a system of non-interacting electrons having a Hamilton operator

\[ \hat{H}_S = \hat{T}_S + \hat{V}_S = \sum_{i=1}^{N} \left[ -\nabla_i^2 + v_S(r_i) \right]. \]  \hspace{1cm} (2.40)

Due to the fact, that the electrons do not interact, the resulting Schrödinger equation can be separated and written in terms of so called Kohn–Sham orbitals:

\[ \left[ -\nabla^2 + v_S(r) \right] \psi_i^{KS} = \epsilon_i \psi_i^{KS} \]  \hspace{1cm} (2.41)

Kohn and Sham now assumed that every system represented by \( \rho_0 \) can be described by a system of non-interacting particles with exactly the same electron density \( \rho_S = \rho_0 \). As a consequence, the Kohn–Sham orbitals describing the system of non-interacting particles can be adopted to describe \( \rho_0 \) as well.

The energy functional \( E[\rho] \) is now represented as

\[ E[\rho] = T_s[\rho] + V_{\text{ext}}[\rho] + J[\rho] + \left( V_{\text{xc}}[\rho] - J[\rho] + T[\rho] - T_s[\rho] \right) \]  \hspace{1cm} (2.42)

\[ := E_{\text{xc}}[\rho] \]

\[ = T_s[\rho] + V[\rho] + J[\rho] + E_{\text{xc}}[\rho]. \]  \hspace{1cm} (2.43)

Here, the kinetic energy functional \( T_s[\rho] \) and the Coloumb functional \( J[\rho] \) are known quantities.
Again, the Euler–Lagrange formalism can be applied to minimize the
energy functional with respect to the electron density, i.e.,

\[
\mu = \frac{\delta T_S[\rho]}{\delta \rho(r)} + v_{\text{ext}}(r) + v_{\text{coul}}(r) + v_{\text{xc}}(r) =: v_{\text{eff}}(r)
\]

with the Coulomb potential

\[
v_{\text{coul}}(r) = \frac{\delta J[\rho]}{\delta \rho(r)} = \int \frac{\rho(r')}{|r - r'|} dr'
\]

and the exchange–correlation potential

\[
v_{\text{xc}}(r) = \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(r)}.
\]

It should be noted that the mathematical form of the Kohn–Sham
equation (2.47) resembles the one of the Hartree–Fock (HF) equation
(2.48):

\[
\left[ -\frac{1}{2} \nabla^2 + v(r) + v_{\text{coul}}(r) + v_{\text{xc}}(r) \right] \psi_{i}^{\text{KS}} = \epsilon_i \psi_{i}^{\text{KS}}
\]

\[
\left[ -\frac{1}{2} \nabla^2 + v(r) + \sum_{j}^{\text{occ}} (\hat{J}_j + \hat{K}_j) \right] \psi_{i}^{\text{HF}} = \epsilon_i \psi_{i}^{\text{HF}}
\]

In essence, these two equations only differ in the sums of exchange
operators and the exchange–correlation potential.

Several approximations to the exchange and correlation functionals
exist and can be classified in different categories: Functionals which only
depend on the electron density \( \rho \) are known as local-density approxima-
tion (LDA) functionals [125], or local density approximation functionals.
If the functional depends on \( |\nabla \rho| \) besides \( \rho \), they are classified as general-
ized gradient approximation functionals [126]. For a detailed overview
of the different categories of functionals, e.g., meta-GGA functionals or
hybrid functionals, we refer the reader to excellent literature about this
topic [127–131].
2.3 Gradient-driven Molecule Construction

The main part of this thesis is devoted to the design of molecular scaffolds to stabilize an (activated) molecular fragment. The chemical embedding should then lead to vanishing forces on both the nuclei of the fragment and of the environment, that is, for every nucleus \( I \) the length of the derivative vector of the electronic energy \( E_{el} \) with respect to the Cartesian coordinates \( \mathbf{r}_I \) of nucleus \( I \) must vanish, i.e.,

\[
|\nabla_I E_{el}| = \sqrt{\left( \frac{\partial E_{el}}{\partial r_{I,x}} \right)^2 + \left( \frac{\partial E_{el}}{\partial r_{I,y}} \right)^2 + \left( \frac{\partial E_{el}}{\partial r_{I,z}} \right)^2} \equiv 0, \tag{2.49}
\]

as otherwise the fragment structure will not correspond to a stationary point on the potential energy surface.

In essence, besides having vanishing lengths of all Cartesian gradient components at each nucleus of the (activated) chemical fragment, also all Cartesian gradient components of the added nuclei need to be zero [132].

For this, a two-step optimization procedure can be applied which is called gradient-driven molecule construction [76, 77]: First, the gradient-reducing environment is represented by a so called jacket potential \( v_{jac} \) mediating all interactions with the chemical fragment. Thereafter, a molecular realization of \( v_{jac} \) has to be found.

To this end, the electronic Hamilton operator for the chemical fragment can be set up as a one-electron operator as follows:

\[
H_{el} = -\frac{1}{2} \sum_{i \in \text{frag}} \Delta_i - \sum_{i,j \in \text{frag}} \frac{Z_i}{r_{ij}} + \sum_{i,j \text{frag}} \frac{1}{r_{ij}} + \sum_{I,J \text{frag}} \frac{Z_I Z_J}{r_{IJ}} + \sum_{i \in \text{frag}} v_{\text{jac}}(i). \tag{2.50}
\]

Here, the indices \( I, J \) as well as \( i \) and \( j \) run over all nuclei and electrons, respectively, of the chemical fragment. The spatial distance between two particles \( i \) and \( j \) is denoted by \( r_{ij} \). The mass and charge of nucleus \( I \) are given by \( M_I \) and \( Z_I \), respectively.

In Eq. (2.50), the jacket potential \( v_{\text{jac}} \) is written as a one-electron operator, however, depending on the particular choice of representation of \( v_{\text{jac}} \), it may also include contributions from electronic kinetic energy operators or nuclear repulsion terms.
The many-electron wave function $\Psi_{\text{el}}$ of the chemical fragment can be approximated by a determinant expansion (see Sec. 2.1.4), leading to a self-consistent-field-type equation for the spin orbitals:

$$\left( \frac{1}{2} \gamma_{ii} \Delta_{\text{frag}} + v_{\text{frag}}(r) + v_{\text{jac}}(r) \right) \phi_i(r) = \varepsilon_i \phi_i(r). \quad (2.51)$$

Here, $\phi_i(r)$ represents the $i$-th orbital with the corresponding orbital energy $\varepsilon_i$. The parameter $\gamma_{ii}$ denotes a generalized occupation number, which has a value of one for single-determinant approximations such as Hartree–Fock or Kohn-Sham DFT. For multi-determinant theories, $\gamma_{ii}$ becomes a real number different from one.

The operator $\frac{1}{2} \gamma_{ii} \Delta_{\text{frag}}$ corresponds to the kinetic energy of the chemical fragment (denoted as 'frag'). All potential energy terms are combined into $v_{\text{frag}}(r)$ to yield

$$v_{\text{frag}}(r) = v^{(\text{nn})}_{\text{frag}}(r) + v^{(\text{ne})}_{\text{frag}}(r) + v^{(\text{ee})}_{\text{frag}}(r) + v^{(\text{xc})}_{\text{frag}}(r). \quad (2.52)$$

Here, the labels (n) and (e) denote the nuclei and electrons, respectively. The pairwise repulsion potential of all electrons $v^{(\text{ee})}_{\text{frag}}(r)$ is nothing else than the Coulomb potential $v^{(\text{coul})}_{\text{frag}}(r)$ and the non-classical term $v^{(\text{xc})}_{\text{frag}}(r)$ represents the exchange–correlation potential.

The jacket potential $v_{\text{jac}}$ can be divided into a potential energy operator for the chemical environment and one for the interaction between the chemical environment and the fragment:

$$v_{\text{jac}}(r) = v_{\text{env}}(r) + v_{\text{int}}(r). \quad (2.53)$$

The environment potential $v_{\text{env}}(r)$ can be expressed as

$$v_{\text{env}}(r) = v^{(\text{ne})}_{\text{env}}(r) + v^{(\text{coul})}_{\text{env}}(r) + v^{(\text{xc})}_{\text{env}}(r) + v^{(\text{nn})}_{\text{env}}(r) + v^{(\text{kin})}_{\text{env}}(r), \quad (2.54)$$

where $v^{(\text{kin})}_{\text{env}}(r)$ represents the kinetic energy contributions of the electrons of the chemical environment.

In an analogous way, the interaction potential (denoted by 'int') between chemical environment and chemical fragment can be split into five parts, i.e.,

$$v_{\text{int}}(r) = v^{(\text{ne})}_{\text{int}}(r) + v^{(\text{coul})}_{\text{int}}(r) + v^{(\text{xc})}_{\text{int}}(r) + v^{(\text{nn})}_{\text{int}}(r) + v^{(\text{kin})}_{\text{int}}(r). \quad (2.55)$$
The term $v^{(ne)}_{\text{int}}(r)$ denotes the interaction of the fragment nuclei with the electrons of the environment and the interaction of fragment electrons with the nuclei of the environment. The Coulomb repulsion of the fragment electrons and the electrons of the environment is expressed by $v^{(coul)}_{\text{int}}(r)$, while the exchange–correlation contribution from both the chemical environment and the chemical fragment as well as non-additive kinetic-energy contributions are given by $v^{(xc)}_{\text{int}}(r)$. The mutual repulsion between the fragment nuclei and the nuclei of the environment is expressed by $v^{(nn)}_{\text{int}}(r)$.

The overall absolute gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ of a chemical fragment can now be defined as the sum of all individual absolute gradients:

$$|\nabla_{\text{frag}} E_{\text{el}}| = \sum_{B_{\text{frag}}} |\nabla B E_{\text{el}}|.$$  \hspace{1cm} (2.56)

The expression $\frac{\partial E_{\text{el}}}{\partial r_{I,\alpha}}$ involved in Eq. (2.49) is comprised of several contributions:

$$\frac{\partial E_{\text{el}}}{\partial r_{I,\alpha}} = \sum_{B_{\text{frag}}} \frac{Z_I Z_B (r_{I,\alpha} - r_{B,\alpha})}{|r_{I} - r_{B}|^3} - \sum_i \gamma_{ii} \left( \frac{\partial \phi_i(r)}{\partial r_{I,\alpha}} \right) \Delta \phi_i(r)$$  \hspace{1cm} (2.57)

$$+ \int \frac{\partial v_{\text{frag}}(r)}{\partial r_{I,\alpha}} \rho(r) \, dr + \int v_{\text{frag}}(r) \frac{\partial \rho(r)}{\partial r_{I,\alpha}} \, dr$$

$$+ \int \frac{\partial v_{\text{jac}}(r)}{\partial r_{I,\alpha}} \rho(r) \, dr + \int v_{\text{jac}}(r) \frac{\partial \rho(r)}{\partial r_{I,\alpha}} \, dr,$$

with $\alpha \in \{x, y, z\}$.

The first two terms in Eq. (2.57) concern the derivatives of the nucleus–nucleus repulsion energy and the electronic kinetic energy of the fragment.

Again, we can divide the potential terms into different contributions, i.e.,

$$\frac{\partial v^{(ne)}_{\text{frag}}(r)}{\partial r_{I,\alpha}} = -\frac{Z_I (r_{I,\alpha} - r_\alpha)}{|r_I - r|^3},$$  \hspace{1cm} (2.58)

and

$$\frac{\partial v^{(coul)}_{\text{frag}}(r)}{\partial r_{I,\alpha}} = \int \int \frac{\partial \rho(r')}{\partial r_{I,\alpha}} \frac{\rho(r')}{|r - r'|} \, dr \, dr'.$$  \hspace{1cm} (2.59)
Depending on the approximation chosen for the exchange–correlation functional and the interaction potentials, the calculation of the derivatives in Eqs. (2.58) and (2.59) is more or less involved.

Assuming a one-determinant approximation to \( \Psi_{el} \) (see Section 2.1.4), the electron density can be calculated as a sum over the absolute squares of all occupied spin orbitals:

\[
\rho(r) = \sum_i |\phi_i(r)|^2. \tag{2.60}
\]

Specifically, for real-valued orbitals, the following expression is valid:

\[
\frac{\partial \rho(r)}{\partial r_{I,\alpha}} = 2 \sum_i \phi_i(r) \frac{\partial \phi_i(r)}{\partial r_{I,\alpha}}. \tag{2.61}
\]

Adopting an expansion of spin orbitals in terms of atom-centered basis functions as described by Eq. (2.29), then yields:

\[
\frac{\partial \phi_i(r)}{\partial r_{I,\alpha}} = \sum_\mu \frac{\partial c_{\mu i}}{\partial r_{I,\alpha}} \chi_{\mu l}(r - r_I) + \sum_\mu c_{\mu i} \frac{\partial \chi_{\mu i}(r - r_I)}{\partial r_{I,\alpha}}. \tag{2.62}
\]

The requirement of vanishing geometry gradients on all nuclei expressed by Eq. (2.49) is equivalent to having vanishing individual Cartesian components. As a result, the right hand side of Eq. (2.57) must be zero, leading to

\[
- \sum_B \frac{Z_I Z_B (r_{I,\alpha} - r_{B,\alpha})}{|r_I - r_B|^3} + \sum_i \gamma_{ii} \left( \frac{\partial \phi_i(r)}{\partial r_{I,\alpha}} |\Delta \phi_i(r)| \right) \\
- \int \frac{\partial v_{\text{frag}}(r)}{\partial r_{I,\alpha}} \rho(r) \, dr - \int v_{\text{frag}}(r) \frac{\partial \rho(r)}{\partial r_{I,\alpha}} \, dr \\
\quad \equiv \int \frac{\partial v_{\text{jac}}(r)}{\partial r_{I,\alpha}} \rho(r) \, dr + \int v_{\text{jac}}(r) \frac{\partial \rho(r)}{\partial r_{I,\alpha}} \, dr \quad \forall I, \alpha. \tag{2.63}
\]

Even though Eq. (2.63) can be employed to determine the jacket potential analytically, in practice we need numerical methods which approximate \( v_{\text{jac}}(r) \) in an iterative fashion.
2.4 Frozen-Density Embedding

The construction of \( u_{\text{Jac}}(r) \) on, e.g., a DFT grid, requires a well-chosen starting guess in order for the optimization procedure to converge to a chemically meaningful solution. The embedding potential obtained from a frozen-density embedding (FDE) calculation could be employed for this purpose. We therefore review the most important concepts of FDE as introduced by Wesolowski and Warshel in 1993 [133, 134].

The electron density \( \rho(r) \), in general referred to as 'supermolecular density', of a system is split into a sum of two fragment densities \( \rho_1(r) \) and \( \rho_2(r) \) as follows:

\[
\rho(r) = \rho_1(r) + \rho_2(r) \quad (2.64)
\]

In analogy to the derivation for Eq. (2.43), the total energy is expressed as a bi-functional of the electron density:

\[
E_{\text{el}}[\rho_1, \rho_2] = T_s[\rho_1, \rho_2] + V_{\text{ex}}[\rho_1, \rho_2] + J[\rho_1, \rho_2] + E_{\text{xc}}[\rho_1, \rho_2]. \quad (2.65)
\]

However, in contrast to the Kohn-Sham equation in Eq. (2.41), the Kohn–Sham orbitals of the two subsystems are not able to represent the kinetic energy bifunctional. Instead, the kinetic energy terms of the non-interacting subsystems as well as a non-additive term are employed to express the kinetic energy bifunctional as follows:

\[
T_s[\rho_1, \rho_2] = T_s[\rho_1] + T_s[\rho_2] + T_s^{\text{nadd}}[\rho_1, \rho_2]. \quad (2.66)
\]

In the same way, the exchange-correlation energy bifunctional is expanded into a sum of exchange-correlation energies of the two subsystems plus a non-additive term [135], that is,

\[
E_{\text{xc}}[\rho_1, \rho_2] = E_{\text{xc}}[\rho_1] + E_{\text{xc}}[\rho_2] + E_{\text{xc}}^{\text{nadd}}[\rho_1, \rho_2]. \quad (2.67)
\]

Now, the electron density of one subsystem, e.g., \( \rho_2(r) \), is kept fixed (denoted as 'frozen density') and the electron density of the other subsystem, here \( \rho_1(r) \), is determined by minimizing the total energy bifunctional \( E_{\text{el}}[\rho_1, \rho_2] \) with respect to \( \rho_1(r) \), denoted as the 'active density'.

Analogous to the derivation for Kohn–Sham DFT (see Section 2.2.2), a non-interacting reference system is defined which is then compared to
the interacting reference system. As a result, a Kohn–Sham-like equation is obtained:

\[ [-\frac{1}{2} \Delta_i + v^{KS}_\text{eff}[\rho_1](r) + v^{\text{emb}}_\text{eff}[\rho_1, \rho_2](r)] \psi_i(r) = \epsilon_i^{(1)} \psi_i(r). \] (2.68)

Here, the effective potential \( v^{\text{eff}} \) is a sum of an effective Kohn–Sham potential \( v^{KS}_\text{eff} \) arising from subsystem \( \rho_1(r) \),

\[ v^{KS}_\text{eff}[\rho_1](r) = \psi^e_1 + \int \frac{\rho_1(r')}{|r - r'|} d r' + \left. \frac{\partial E_{\text{xc}}[\rho]}{\partial \rho} \right|_{\rho = \rho_1(r)} , \] (2.69)

and of a so-called embedding potential \( v^{\text{emb}}_\text{eff} \) [136] treating the interaction with the frozen density \( \rho_2(r) \),

\[ v^{\text{emb}}_\text{eff}[\rho_1, \rho_2](r) = \psi^e_2(r) + \int \frac{\rho_2(r')}{|r - r'|} d r' \]
\[ + \frac{\delta E_{\text{xc}}^{\text{nadd}}[\rho_1, \rho_2]}{\delta \rho_1} + \frac{\delta T_s^{\text{nadd}}[\rho_1, \rho_2]}{\delta \rho_1} . \] (2.70)

### 2.5 Vibrational Raman Optical Activity Spectroscopy

Vibrational Raman Optical Activity (VROA) measures the difference of the Raman scattering intensity of right and left circularly polarized light. The calculation of VROA of a molecular system requires linear response functions involving the electric dipole, magnetic dipole, and electric quadrupole moment operators which are given by [84, 89, 137]

\[ \hat{\mu}_\alpha = -\sum_i r_{i\alpha}, \] (2.71)

\[ \hat{m}_\alpha = -\sum_i \frac{1}{2} l_{i\alpha} = -\sum_{i,\alpha,\beta} \frac{1}{2} \epsilon_{\alpha\beta\gamma} r_{i\beta} p_{i\gamma}, \] (2.72)
and

\[ \hat{\Theta}_{\alpha\beta} = -\frac{1}{2} \sum_i \left( 3r_{i\alpha} r_{i\beta} - \delta_{\alpha\beta} r_i^2 \right), \]  

(2.73)

respectively. The summations run over the number of electrons, and the length representation of the electric dipole and quadrupole operators is employed. The position and momentum operators of electron \( i \) are given by \( r_i \) with Cartesian components \( r_{i\alpha}, r_{i\beta}, r_{i\gamma} \), and by \( p_i \) with its components \( p_{i\alpha}, p_{i\beta}, p_{i\gamma} \), respectively. Further, \( \epsilon_{\alpha\beta\gamma} \) is the third-rank antisymmetric tensor. In SI-based atomic units, the nonrelativistic form of the magnetic moment operator is given in terms of the orbital angular momentum \( l_{i\alpha} \) by Eq. (2.72). The components of the complex electric dipole-dipole polarizability \( \alpha_{\alpha\beta} \), the electric dipole-magnetic dipole polarizability \( G'_{\alpha\beta} \), and the electric dipole-quadrupole polarizability \( A_{\gamma\delta\beta} \) tensors are given by

\[ \alpha_{\alpha\beta} = \alpha_{\alpha\beta}^R + i\alpha_{\alpha\beta}^I, \]  

(2.74)

\[ G'_{\alpha\beta} = G'_{\alpha\beta}^R + iG'_{\alpha\beta}^I, \]  

(2.75)

and

\[ A_{\alpha\beta\gamma} = A_{\alpha\beta\gamma}^R + iA_{\alpha\beta\gamma}^I, \]  

(2.76)

in terms of their real \((R)\) and imaginary \((I)\) parts. Adopting a sum-over-states (SOS) formulation, for convenience of notation, the real and imaginary components of the polarizability tensors can be defined in terms of the multipole operators of Eqs. (2.71) – (2.73) as

\[ \alpha_{\alpha\beta}^R = \sum_{n=0}^f (w_{n0}, \omega, \Gamma) \text{Re} \left[ \langle \Psi_0 | \hat{\mu}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{\mu}_\beta | \Psi_0 \rangle \right], \]  

(2.77)

\[ \alpha_{\alpha\beta}^I = \sum_{n=0}^f (w_{n0}, \omega, \Gamma) \text{Re} \left[ \langle \Psi_0 | \hat{\mu}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{\mu}_\beta | \Psi_0 \rangle \right], \]  

(2.78)

\[ G'_{\alpha\beta}^R = \sum_{n=0}^f (w_{n0}, \omega, \Gamma) \text{Re} \left[ \langle \Psi_0 | \hat{\mu}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{\mu}_\beta | \Psi_0 \rangle \right], \]  

(2.79)

\[ G'_{\alpha\beta}^I = \sum_{n=0}^f (w_{n0}, \omega, \Gamma) \text{Re} \left[ \langle \Psi_0 | \hat{\mu}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{\mu}_\beta | \Psi_0 \rangle \right], \]  

(2.80)

\[ A_{\alpha\beta\gamma}^R = \sum_{n=0}^f (w_{n0}, \omega, \Gamma) \text{Re} \left[ \langle \Psi_0 | \hat{\mu}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{\Theta}_{\gamma\beta} | \Psi_0 \rangle \right], \]  

(2.81)

\[ A_{\alpha\beta\gamma}^I = \sum_{n=0}^f (w_{n0}, \omega, \Gamma) \text{Re} \left[ \langle \Psi_0 | \hat{\mu}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{\Theta}_{\gamma\beta} | \Psi_0 \rangle \right]. \]  

(2.82)
The dispersion line-shape functions \( f^\pm(\omega_{n0}, \omega, \Gamma) \) and absorption line-shape functions \( g^\pm(\omega_{n0}, \omega, \Gamma) \) related to finite life time broadening are given by [138, 139]

\[
f^\pm(\omega_{n0}, \omega, \Gamma) = \frac{\omega_{n0} - \omega}{(\omega_{n0} - \omega)^2 + \Gamma^2} \pm \frac{\omega_{n0} + \omega}{(\omega_{n0} + \omega)^2 + \Gamma^2}, \tag{2.83}
\]

\[
g^\pm(\omega_{n0}, \omega, \Gamma) = \frac{\Gamma}{(\omega_{n0} - \omega)^2 + \Gamma^2} \pm \frac{\Gamma}{(\omega_{n0} + \omega)^2 + \Gamma^2}. \tag{2.84}
\]

To account for the broadening of the excited electronic states in the calculation of the polarizability tensors, a damping parameter \( \Gamma \) is included phenomenologically [138, 140–142] describing the relaxation and dephasing of the excited states (see also Refs. [143, 144]). In the equations above, \( \Psi_0 \) is the electronic ground state wave function and \( \Psi_n \) is the \( n \)-th excited state wave function. Further, \( \omega \) is the angular frequency of the incident light, \( \hbar \omega_{n0} = E_n - E_0 \) the excitation energy and \( 1/\Gamma \) the life time of the excitation which equals the full-width at half maximum (FWHM) of the absorption band. In the SOS formulation, the broadening \( \Gamma \) may be different for different electronic states. The KS-based linear response implementation employed in this thesis is based on a common broadening parameter, as in the approach of Jensen et al. [145]. Furthermore, the following equations hold for Gaussian and Lorentz broadening [146], respectively:

\[
G(x, x_0, \sigma) = \frac{1}{\sqrt{2 \times \pi \times \sigma}} \times e^{-(x-x_0)^2/(2 \times \sigma^2)}, \tag{2.85}
\]

and

\[
L(x, x_0, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(x - x_0)^2 + (\frac{\Gamma}{2})^2}. \tag{2.86}
\]

For a given one-particle basis set, the contributions from all electronic singlet excited states that can be formed contribute implicitly to the response functions; that is, these are formally equivalent to a full SOS within a given basis. Only states with relatively low energy correspond to physical electronic excited states of the system, but the full set is needed in order to exploit the full flexibility of the one-particle basis in the response calculations. Further, the usual approximations of KS theory apply, namely the use of approximate functionals for the ground state and approximate adiabatic linear response kernels and the assumption that the
Chapter 2  |  THEORY

perturbed KS density matrix not only provides accurate electric but also magnetic-electric polarizabilities.

For theoretical VROA spectroscopy, the Placzek approximation [147, 148] is adopted to calculate the polarizability transition tensors from the corresponding molecular property tensors. This theory can be extended to a Placzek-like polarizability theory [149] for both on- and off-resonance cases. Following this approach, the transition polarizability tensors are expanded into a Taylor series around the equilibrium geometry. In each expansion, the first-order term accounts for Rayleigh scattering, while the second-order term accounts for Raman/VROA scattering when adopting the harmonic approximation (see Section 2.1.3). Taking the derivative of the molecular property tensors with respect to the normal mode $Q_p$ of the $p$-th vibration, those second terms are defined as follows:

$$\alpha_{\alpha\beta}^P \alpha_{\alpha\beta}^P = \langle 0|\alpha_{\alpha\beta}|1_p\rangle \langle 1_p|\alpha_{\alpha\beta}|0\rangle = \left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_p}\right)_0 \left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_p}\right)_0,$$  \hspace{1cm} (2.87)

$$\alpha_{\alpha\beta}^P G_{\alpha\beta}^{\prime P} = \langle 0|\alpha_{\alpha\beta}|1_p\rangle \langle 1_p|G_{\alpha\beta}^{\prime}|0\rangle = \left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_p}\right)_0 \left(\frac{\partial G_{\alpha\beta}^{\prime}}{\partial Q_p}\right)_0,$$  \hspace{1cm} (2.88)

and

$$\alpha_{\alpha\beta}^P \epsilon_{\alpha\gamma\delta} A_{\gamma,\delta\beta}^{\prime P} = \langle 0|\alpha_{\alpha\beta}|1_p\rangle \langle 1_p|\epsilon_{\alpha\gamma\delta} A_{\gamma,\delta\beta}^{\prime}|0\rangle = \left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_p}\right)_0 \epsilon_{\alpha\gamma\delta} \left(\frac{\partial A_{\gamma,\delta\beta}^{\prime}}{\partial Q_p}\right)_0.$$

These derivatives may be calculated by an $n$-point central difference formula [150] by performing $n-1$ distortions of every Cartesian nuclear coordinate of the molecule.
Vibrational Raman Optical Activity Spectroscopy  |  2.5

The transition polarizability tensors are then contracted to different isotropic and anisotropic invariants:

\[
\alpha_p^2 = \text{Re} \left( \frac{1}{9} \alpha_{a\alpha}^p \alpha_{\beta\beta}^{p\ast} \right), \quad (2.90)
\]

\[
\beta(\alpha)_p^2 = \text{Re} \left( \frac{3 \alpha_{a\beta}^p \alpha_{\alpha\beta}^{p\ast} - \alpha_{a\alpha}^p \alpha_{\beta\beta}^{p\ast}}{2} \right), \quad (2.91)
\]

\[
\beta(G')_p^2 = \text{Im} \left( \frac{3 \alpha_{a\beta}^p G_{a\beta}^{p\ast} - \alpha_{a\alpha}^p G_{\beta\beta}^{p\ast}}{2} \right), \quad (2.92)
\]

\[
\beta(A)_p^2 = \text{Re} \left( \frac{1}{2} \omega \alpha_{a\beta}^p \epsilon_{ay\delta} A_{y,\delta\beta}^{p\ast} \right), \quad (2.93)
\]

\[
(\alpha G')_p^2 = \text{Im} \left( i \frac{1}{9} \alpha_{a\alpha}^p G_{\beta\beta}^{p\ast} \right), \quad (2.94)
\]

where \(\beta(\times)_p^2\) is the anisotropic invariant of the transition tensor '\(\times\)', and \(\text{Re}(\times)\) and \(\text{Im}(\times)\) denote the real and imaginary components of the invariants, respectively. In all equations above, the Einstein summation convention is applied for repeated Greek indices. A Greek subscript denotes either \(x\), \(y\), or \(z\) in Cartesian coordinates.

Substituting the complex expressions for the different property tensors given by Eqs. (2.74), (2.75), and (2.76) yields

\[
\alpha_p^2 = \frac{1}{9} \left( \alpha_{a\alpha}^{R,p} \alpha_{\beta\beta}^{R,p} + \alpha_{a\alpha}^{I,p} \alpha_{\beta\beta}^{I,p} \right), \quad (2.95)
\]

\[
\beta(\alpha)_p^2 = \frac{1}{2} \left( 3 \left( \alpha_{a\beta}^{R,p} \alpha_{a\beta}^{R,p} + \alpha_{a\beta}^{I,p} \alpha_{a\beta}^{I,p} \right) - \alpha_{a\alpha}^{R,p} \alpha_{\beta\beta}^{R,p} - \alpha_{a\alpha}^{I,p} \alpha_{\beta\beta}^{I,p} \right), \quad (2.96)
\]

\[
\beta(G')_p^2 = \frac{1}{2} \left( 3 \left( \alpha_{a\beta}^{R,p} G_{a\beta}^{R,p} + \alpha_{a\beta}^{I,p} G_{a\beta}^{I,p} \right) - \alpha_{a\alpha}^{R,p} G_{\beta\beta}^{R,p} - \alpha_{a\alpha}^{I,p} G_{\beta\beta}^{I,p} \right), \quad (2.97)
\]

\[
\beta(A)_p^2 = \frac{1}{2} \left( \omega \alpha_{a\beta}^{R,p} \epsilon_{ay\delta} A_{y,\delta\beta}^{R,p} + \omega \alpha_{a\beta}^{I,p} \epsilon_{ay\delta} A_{y,\delta\beta}^{I,p} \right), \quad (2.98)
\]

\[
(\alpha G')_p^2 = \frac{1}{9} \left( \alpha_{a\alpha}^{R,p} G_{\beta\beta}^{R,p} + \alpha_{a\alpha}^{I,p} G_{\beta\beta}^{I,p} \right). \quad (2.99)
\]

For off-resonance VROA spectroscopy, the imaginary components of the different property tensors in the equations above are zero. Yet, if the excitation wavelength \(\lambda\) is close in energy to an electronic transition of the studied molecular system, the imaginary components of the property
tensors are of comparable magnitude to their real counterparts and must be considered.

In practice, VROA measurements can be performed employing different experimental setups, each with a different relation between the molecular properties and the measured intensities. In this thesis, we choose a backscattering geometry where the VROA intensities are given by [151]

\[ I^R (180°) - I^L (180°) = \Delta \frac{d\sigma}{d\Omega} (180°) \propto \frac{4}{c} \left( 24\beta(G')^2 + 8\beta(A)^2 \right) \]  \hspace{1cm} (2.100)

Here, \( c \) is the speed of light, \( \beta(G')^2 \) is the anisotropic invariant of the product of the electric dipole-dipole polarizability transition tensor with the electric dipole-magnetic dipole polarizability transition tensor, and \( \beta(A)^2 \) is the anisotropic invariant of the product of the electric dipole-dipole polarizability transition tensor with the electric dipole-electric quadrupole polarizability transition tensor.
Molecular Scaffold Design by Gradient-driven Molecule Construction

Reactive molecules usually represent activated species that are decisive for accomplishing a chemical transformation. Prominent examples are transition states, exotic molecules (such as multiply bonded transition metal dimers, small carbon cages, strained ring structures, and radicals [152–157]), and species that activate inert molecules upon binding. A suitable molecular environment may stabilize such peculiar local bonding situations. For example, cyclobutadiene is very unstable in its free form [158], but it exists in its planar, rectangular structure in some metal compounds [159].

Identifying such molecules with desired properties is at the heart of rational compound design. In computational approaches, this is generally accomplished by high-throughput screening. Recent examples in the field of materials science comprise national initiatives such as the materials project [160] and the MARVEL national center of competence in research [161], but also local efforts such as the Harvard clean energy project [162], the screening for electrolyte solvents by Korth and co-workers [163], the search for insulation gases by Franck [164], and the extensive efforts to exploit machine learning by Lilienfeld and co-workers [165–169].

Even inverse approaches have been considered in order to directly exploit the ill-defined direction from desired property to structure [170]. To this end, several automated schemes have recently been proposed
for the automated construction of stabilizing molecular environments leading to a desired property. Avendaño-Franco et al. [171] introduced the firefly algorithm as a means to discover a variety of novel competitive metastable structures. Kulik and collaborators [172] presented a toolkit for first-principles screening and discovery of new inorganic molecules. For further examples, we direct the reader to the literature where some excellent reviews and references to this topic exist [76, 169, 170, 173–175]. Although intellectually appealing, such inverse approaches are still underdeveloped and cannot compete with high-throughput screening yet. In particular, since the Hamilton operator can possess similar eigenvalues for completely different molecular compositions, inverting the Schrödinger equation is ill-defined [176], even with a fixed eigenvalue spectrum [177]. Therefore, new concepts in inverse design are required which do not try to directly invert the Schrödinger equation, while at the same time, they reduce the complexity of a screening approach. After all, the nearly infinite amount of organic molecular structures ($10^{20}$ to $10^{24}$ molecules) accessible by current synthetic protocols, renders any high-throughput screening for a specific desired property nearly impossible [178, 179].

In this chapter, we develop a sequential and automated scaffold construction approach that aims at stabilizing “activated” molecular fragments. We first explain how such an environment could in principle be generated by an analytic environment potential, before we turn to building up the chemical environment atomwise in an iterative fashion. An activated fragment shall be a molecular substructure in an otherwise stable molecule that would not be stable when isolated. Typically, such a fragment structure will feature structural distortions compared to its fully relaxed isolated parent. Clearly, also reactive species or transition state structures can serve as an activated fragment (in fact, theozyme concept [51, 180, 181] could be subjected to our embedding construction). It is then the role of the environment to stabilize such species in such a way that a (super)molecule results that is a stationary point on a potential energy hypersurface.

We consider the “activated fragment” a concept to design catalytic pathways, although it is not limited to this type of application. For a catalytic pathway, any reactant or substrate requires at some point bond activation by structural distortion to lower activation barriers. The distorted structure will require some sort of stabilization through chemical bonding to the catalyst, which is the target of our design protocol. Whereas we
develop our tool at the example of a stable structure, we may as well
define the design goal as a transition state with a bound on its energy
compared to some minimum structure (i.e., a bound on the activation
energy). Instead of applying a high-throughput screening approach of
existing scaffolds, we aim at the construction of a shell-structured envi-
ronment tailored to increase the stability of the activated fragment by a
growing number of (atomic) onion shells surrounding the fragment to
be stabilized. This procedure shall ensure that a greater variety of struc-
tures can be explored and that unusual embedding environments can be
discovered.

The automated atom-wise ligand environment construction outlined
in this chapter is an efficient tool to generate complex molecular scaffolds.
They are hardly accessible by manual inspection because of their sheer
number. This number is a result of the combinatorial explosion of possible
structures, which is a sign of the vast chemical compound space. A design
study now attempts to fish in this sea of structures for those that show
certain properties. For one such property, only a very limited number of
structures will be relevant. For the design process to be feasible at all, it is
decisive to limit and focus the structure search. A branching out scaffold
construction will have a vanishing effect on a local property located in its
center. In principle, any scoring function can then be applied to prune
the tree of structures. Here, we choose GdMC in combination with ligand
binding energies, but we emphasize that any other scoring function can
be chosen.

We choose GdMC because its ingredients, the nuclear gradients, are a
local property, which is mostly defined by the close neighborhood of the
activated fragment. However, the activated fragment itself serves another
purpose in GdMC. By construction it is not relaxed to study the (sought-
for diminishing) effect on the nuclear gradients. As a consequence, all
ligand structures constructed are directly comparable because they were
obtained for the same frozen fragment (by contrast to straightforward
screening approaches). If all structures were to be fully relaxed, the nature
of the fragment would be changed and no comparative conclusion could
be drawn. Hence, although full structural relaxation will eventually be
necessary, it would not serve the proper design target. For instance, if we
search for a complex that produces a diazenoid $N_2$ structure on binding of
$N_2$, we will start from a metal atom binding dinitrogen in which the $N$–$N$
bond length is set to that of diazene and the two nitrogen atoms form an
angle of about 120 degrees with the metal atom. If different ligand environments are combined with such a fragment, full structure optimization will produce, in most cases, linearly bound and hardly activated N$_2$ or a dissociation of the ligand, which clearly is not the design target. Moreover, little can then be learned about why the diazenoid structure cannot be stabilized or how it could be stabilized.

By freezing the activated fragment in terms of geometry constraints, we define in advance what type of fragment geometry we want the target complex to stabilize. Since all investigated structures then have the same geometrical constraints on the fragment, we are able to directly compare them with each other in terms of their stabilizing effect of their ligand environments on the activated fragments. As a result, the GdMC concept defines a way to compare different ligand environments based on their effect on the same target structure; this is illustrated in Fig. 3.1. Clearly, full structure optimization would be employed once a suitable candidate structure was identified by GdMC.
Figure 3.2.: The dihydrogen molecule, together with the oxygen atom that stabilizes the enlarged bond $d_{\text{H}_2\text{O}}$, forming a water molecule. The goal is now to represent the environment potential of the oxygen atom by means of an embedding potential based on an FDE starting guess. This figure was reprinted from Ref. [184].

3.1 Generating a Molecular Environment by Frozen-Density Embedding

As a means to show how the analytic construction of a chemical environment can be accomplished, we study the case of a dihydrogen molecule with an elongated bond corresponding to the H···H distance in a water molecule (see Fig. 3.2 for an illustration). The first pilot study [77] aimed at finding an environment potential $v_{\text{jac}}$ on a DFT grid which stabilizes this H···H fragment, i.e., ideally recovering the chemical environment which could be represented by a sole oxygen atom. As an initial guess for such a jacket potential which is added to the Kohn–Sham potential, a zero vector was fed into a modified version of TURBOMOLE 5.7.1 [182]. Applying the Nelder–Mead simplex algorithm [183] allowed for the minimization of the overall gradient on the chemical fragment, with the number of variational parameters purely determined by the DFT gridsize. It was shown that the overall fragment gradient could be significantly reduced, thus demonstrating the applicability of this approach. However, it was pointed out that the reconstructed potential does not show $C_{2v}$ symmetry and could, in general, not be represented by a chemical structure, that is, the potential is not $C$-representable. This in turn raises the question how much a-priori information about the chemical environment has to be provided such that a $C$-representable environment potential will be the natural outcome of such an approach.

As a step towards improving this methodology, we propose a starting guess for the jacket potential $v_{\text{jac}}$ based on a FDE calculation. Comparing the equation for the embedding potential,
\[ \nu_{\text{emb}}[\rho_1, \rho_2](r) = \nu_{2k}^\text{el}(r) + \int \frac{\rho_2(r')}{|r - r'|} dr' + \delta E_{xc}^{\text{nadd}}[\rho_1, \rho_2] \delta \rho_1 \delta \rho_1 + \delta T_s^{\text{nadd}}[\rho_1, \rho_2] \delta \rho_1 \]

with the defining equation for the jacket potential,

\[ \nu_{\text{jac}}(r) = \nu_{\text{env}}(r) + \nu_{\text{int}}(r). \]

leads to the following conclusion:

\[ \nu_{\text{eff}}^{\text{emb}} \approx \nu_{\text{jac}}. \]

We note that Eq. (3.3) merely suggests a correspondence between \( \nu_{\text{eff}}^{\text{emb}} \) and \( \nu_{\text{jac}} \) as the studied system \( d_{H_2O} \) is characterized by strongly covalent O···H bonds. In essence, the activated H···H system does not know about these interactions beforehand and the cut across these bonds will affect their correlation energy [185]. Still, it is worthwhile to employ \( \nu_{\text{eff}}^{\text{emb}} \) as a starting guess for the optimization of the jacket potential, that is,

\[ \min_{\nu_{\text{jac}}} |\nabla_{\text{frag}} E_{\text{el}}(\nu_{\text{jac}})|. \]

### 3.1.1 Challenges with the FDE Starting Guess

The first test calculations employing \( \nu_{\text{eff}}^{\text{emb}} \) of the oxygen atom as a starting guess (see Fig. 3.3) for \( \nu_{\text{jac}} \) did not accelerate nor improve the overall reconstruction procedure. Several possible reasons for this could be identified, one of them being the distribution of electrons as well as the choice of the electronic configuration of the chemical fragment. Particularly in the case of the dihydrogen molecule, two possible electronic configurations could generally be chosen (see Fig. 3.4), i.e., a singlet or triplet configuration. While the singlet state for the H···H at equilibrium distance is clearly favored over the triplet configuration, this is not so obvious for the dihydrogen fragment in the water molecule without the oxygen atom. Here, the energy difference between both configuration is about 0.1 Hartree. As a result, the triplet configuration as a starting configuration should in
Figure 3.3.: **Left:** Radial plot for different potentials of the H$_2$O molecule in the singlet configuration with the oxygen atom at the coordinate origin and the hydrogen atoms at around 1.0 Ångstrom. The embedding potential $\nu_{\text{emb}}^\text{eff}$ is shown in blue color, the effective KS potential $\nu_{\text{eff}}^\text{emb}$ of the oxygen atom in yellow color, and the effective KS potential $\nu_{\text{eff}}^\text{emb}$ of the dihydrogen fragment in green color. **Right:** The embedding potential $\nu_{\text{eff}}^\text{emb}$ on a DFT grid including a ghost atom at the position of the oxygen atom. Both figures were reprinted from Ref. [184].

Figure 3.4.: **Left:** Dissociation curve of the dihydrogen molecule. Here, $r_{12}$ corresponds to the distance between both hydrogen atoms, whereas $r_{\text{eq}}$ corresponds to the equilibrium distance between both hydrogen atoms in the water molecule. **Right:** Dissociation curve for the H$_2$O molecule. Here, $r_{12}$ is varied with respect to $r_{\text{eq}}$. These figures were reprinted from Ref. [184].
principle be tested for the dihydrogen fragment within the water molecule as well.

Furthermore, the representation of an embedding environment by a radial DFT grid (as is the case with TURBOMOLE), might not be the best option for the design approach at hand. While the numeric integration of atom-centered potentials can be carried out very efficiently on such grids, important grid points, i.e., basis functions, are missing in the region of the environment potential with important structural features. Here, both hydrogen atoms are well represented by a large number of grid points (see Fig. 3.5), which is not the case for the FDE embedding potential of the oxygen atom. As a workaround, basis functions could be introduced into the DFT grid by means of so called ghost atoms (see the right hand side of Fig. 3.5). This would introduce additional constraints about the location of a to-be-constructed atom represented by \( v_{\text{eff}} \) before the start of the design process and it would be desirable not to restrict the optimization procedure in such a way already at the beginning.

Finally, we note that the Nelder–Mead simplex algorithm only accounts for the numerical values of a function during the optimization process, i.e., the numerical value of the environment potential at every grid point. Clearly, it should be the goal to improve this optimization procedure by including gradient information of the FDE embedding potential as well.
[186], which in turn would make other optimization algorithms, such as the conjugate gradient method [187, 188], accessible for this task.

In addition, the quality of the FDE initial starting guess needs to be assessed more thoroughly. Here, a quality measure could be employed, such as the integral of the norm of the displacement density \( d \) which indicates how many electrons are shifted by FDE, i.e.,

\[
d = \frac{1}{2N} \int |\rho^{\text{DFT}} - \rho^{\text{FDE}}| \, dr
\]

Further investigations with regards to this topic are beyond the scope of this thesis, however, work along those lines is currently carried out in our laboratory.

### 3.2 Constructing the Chemical Environment by Atomwise Additions

Instead of approximating \( v_{\text{jac}} \) by an analytic chemical potential, we now turn to the atomwise construction of the molecular environment and introduce a sequential and automated scaffold construction approach that aims at stabilizing activated molecular fragments. Instead of applying a high-throughput screening approach of existing scaffolds, we aim at the construction of a shell-structured environment tailored to increase the stability of the activated fragment by a growing number of (atomic) onion shells surrounding the fragment to be stabilized. As a scoring function, we exploit the concept of GdMC [76, 77, 132] introduced in Chapter 2.3, which we fully automate for this study.

One option for the environment construction is a systematic build-up of the scaffold structure, as proposed recently by Jensen and collaborators [189], which could be exploited here as well. In a proof-of-principle study [77], however, our group manually constructed a chelate ligand for stabilizing bound dinitrogen by subsequently binding individual atoms around a predefined metal fragment. This atomwise construction is advantageous as it allows us to produce unusual chelate ligand environments with special properties that might otherwise be overlooked in a fragment based approach. However, a fully automatized framework is required in order to carry out such a scaffold design in practice. In the subsequent sections, we introduce a fully automatized algorithm called 'Molecular Scaffold
Designer’ (MSD). We apply our framework to consider the design of a peculiar ruthenium complex which is known to bind inert CO₂.

3.3 The Design Target: CO₂ Activation

Carbon dioxide hydrogenation has been the focus of several computational investigations because of its potential importance for clean-energy and climate-protection technologies. For example, Mondal et al. [190] designed a series of potential catalysts and computationally investigated their reactivity towards hydrogenation of CO₂. Ramakrishnan et al. [191] modelled CO₂ insertion free energies and pathways and made attempts to design a reversible, ergoneutral ligand for conversion of CO₂. Several ruthenium-based complexes are known [192–194] which stabilize the CO₂ molecule in a specific geometry.

Fig. 3.6 presents one example that contains a bicyclic pyridinylazo-lato ligand together with two phosphine ligands and one hydrogen atom. Derivatives of this complex in which a hydrogen atom at the pyridinyl moiety is substituted by a methyl or phosphine group were studied with respect to their catalytic activity for carbon dioxide hydrogenation [195]. For establishing an automated, shell-wise ligand construction algorithm with GdMC as a scoring principle, we will design this CO₂-binding complex from scratch. For the implementation and its analysis, our design strategy is such that we will find the compound shown in Fig. 3.6 among other CO₂-fixating complexes constructed. From all these candidate structures, we aim to understand the peculiar structure of the known complex in Fig. 3.6, in particular the importance of the pyridinylazolato ligand. We note that the GdMC principle may be easily exchanged or supplemented by other scoring functions if deemed suitable.

3.4 Protocol for Shell-wise Construction of an Embedding Environment

Without the embedding environment of a chelate ligand, a reactive fragment such as the bent CO₂ ligand in Fig. 3.6 would be structurally unstable. This instability is characterized by non-vanishing nuclear gradients on all of its atoms as the fragment structure will not correspond to a minimum
on the potential energy hypersurface. While several methods exist which build up a molecular environment by placing predefined ligands around a molecular fragment [196–199], the MSD [76, 77, 132] framework builds up the surrounding scaffold structure atom-wise in an iterative fashion producing vanishing nuclear gradients on all atoms. The main advantage of this approach is the full flexibility when constructing the chemical environment in the vicinity of the CO₂ ligand.

### 3.4.1 Binding Sites and Coordination Geometry

As a first step of our heuristics-guided approach, we define binding sites around the activated fragment. Depending on the desired coordination polyhedron of the scaffold to be constructed, the number of binding sites as well as their position relative to the activated fragment are set. For a graphical explanation, see Fig. 3.7.

In the second step, atoms are placed around the activated fragment at the position of the binding sites. Based on the number and type of atoms and on the number of available binding sites, the number of newly generated structures may already grow exponentially. For our example of stabilizing a bent CO₂ moiety bound to a Ru atom, we start with five binding sites and five different types of atoms (H, C, N, O, P). Naturally, each atom type can be placed at each binding site so that all possible permutations have to be accounted for. Note that some combinations of atoms may form stereoisomers. Furthermore, to generate chemically
Figure 3.7.: The activated fragment (here, a bent CO$_2$ molecule coordinated to a ruthenium atom) surrounded by a predefined number of binding sites (here: 3 (left), 4 (middle), and 5 (right)) which determines the overall coordination number of the embedding environment to be constructed.

meaningful structures, hydrogen atoms are attached to each of the newly positioned atoms (allowing for unsaturated structure elements). For instance, at a nitrogen atom, two or three hydrogen atoms may be added leading to NH$_2$ and NH$_3$ groups, respectively. Note that even for a hydrogen atom, we may consider adding another hydrogen atom to produce an H$_2$ ligand.

For the combination of binding sites and atom types given above, $12 	imes 389$ structures can be generated which are unique with respect to their orientation in three dimensional space. This number already occurs in the first iteration step, i.e., for the first generation.

3.4.2 Scoring and Constrained Optimization

In a subsequent step, the newly generated structures are then optimized, keeping the activated fragment (the metal center together with the coordinating CO$_2$ in our example) frozen. The target geometry of the activated fragment is directly encoded in the optimization procedure by positional constraints in terms of internal coordinates for the activated fragment. The value for the nuclear gradient on all atoms of the activated fragment then indicates whether the constructed embedding environment is able to stabilize the (frozen) reactive fragment.

After evaluating the nuclear gradient of the fragment for all optimized structures, a set of suitable candidate structures for the next generation is selected. For this and the following substitution steps, the hydrogen atoms in the outermost shell form the substitution sites and are replaced by the predefined set of atom types employed for the construction. In principle,
the atom types for each substitution step can be chosen individually. If one chooses to restrict the chemical compound space to be screened, the size of the atom set to be substituted in subsequent generations may be smaller.

The newly formed structures are then constrained-optimized again as in the previous step and the procedure starts again. This repetition of steps is then carried out until one or more distinct embedding environments around the activated fragment have been identified as candidates leading to a vanishing nuclear gradient for the activated fragment. Clearly, this procedure does not guarantee that experimentally stable or easily accessible ligand structures are constructed. The purpose of the procedure is to understand what atom arrangements may reduce the forces on all atoms of an activated fragment.

### 3.4.3 Multidentate Ligands and Ring Structures

Crucial for the formation of successful candidates for a stabilizing chemical environment can be the introduction of multidentate ligands, haptic ligands (i.e., the direct coordination of a ligand to the transition metal center by means of a contiguous set of atoms) and cyclic structural features such as (non)aromatic ring structures. Particularly the introduction of multidentate ligands leads to a stabilization of the scaffold around the transition metal center due to the chelate effect [200]. In recent work, Jensen and collaborators [201–204] emphasized the necessity of including chelate ligands in their molecular design approaches.

As the chelate effect on the overall stability of a complex is an entropy effect driven by the release of monodentate ligands from the binding sites then occupied by the multidentate ligand, it is not covered by our design process that relies on electronic energy considerations of only a single species [205]. Still, for our purpose, it is desirable to consider all different types of chelate ligands as early as possible in the construction process in order to investigate their potential for the reduction of forces on all atoms. As a side effect, we may safely assume that chelate structures will eventually lead to complexes more stable against decomposition than those solely built from monodentate ligands [206, 207].

However, depending on the number of binding sites, the variable bridge length of multidentate ligands, and the possibility of self-assembling monodentate ligands [208], the number of feasible scaffolds will dramatically
increase, contributing to the combinatorial explosion of possible candidate structures. Hence, it will be essential to rigorously catalog how each structure was generated in order to trace back a possible candidate structure to its structural ancestor and to identify structural features as early as possible, if they have the desired effect on the nuclear gradient of the activated fragment.

In many cases, one parent structure may be able to form more than one ring topology. Then, MSD ensures that all combinations are built. For example, if the algorithm detects two possible ring prospects, e.g., a five-membered and a six-membered ring, three distinct structures will be generated. That is, one with only a five-membered ring, the second with only a six-membered ring, and, if possible, a third structure containing both ring topologies. Naturally, the unsaturated and aromatic counterparts are also generated, whenever possible.

### 3.4.4 Structure Tree and Structural Inheritance

For the following, we identify the *structure tree* as the collection of all generated structures, each being a member of a *structure branch*. This concept is illustrated on the left in Fig. 3.8. The activated fragment together with the predefined binding sites is identified as the *base structure* from which a set of *child structures* is generated. In subsequent substitutions, the child structures of generation (n-1) form *parent structures* for the n-th substitution from which a new set of child structures is generated. The ordered sum of all *edges* leading from the base structure to one child structure then forms the structure branch. This concept allows us to readily identify each structure and its structural ancestors.

We therefore write \([a \ b \ c \ d \ldots]_{\text{frag}}^{n_{bs}}\) for a structure branch originating from the activated fragment (frag) and the number of binding sites \(n_{bs}\). The number of identifiers (a, b, c, d, ...) indicates the length of the structure branch (e.g., four identifiers represent four generations, that is, four iterative additions of atoms in the example above). Structure d is formed out of its parent structure c which in turn originates from structure b and so forth. Each identifier is a number assigned to a structure within one generation. If there are 500 structures in the first generation, then ‘a’ will run from 1 to 500. If now the second generation contains 1000 child structures originating from one parent structure, then ‘b’ will run from 1
to 1000. For each individual structure branch and each new generation, we start the structure count from 1.

If a chelate or ring topology can be constructed in the n-th generation, we will add an additional index to the identifier of this generation. A chelate ligand is indicated by $c^y_{\tilde{x}}$ where the vector $\tilde{x}$ gives the number of bridges and $y$ the corresponding bridge lengths. For a non-aromatic ring system, we write $r^y_{\tilde{x}}$ with $\tilde{x}$ containing the number of rings with the corresponding number of atoms given by $y$. For aromatic ring systems, the same convention holds with $a^y_{\tilde{x}}$. For example, the term $[9 186c^4_1 765c^4_{2,1}^5 976a^6]_{\text{CO}_2}$ represents a structure branch with the introduction of a chelate ligand of bridge length 4 in the second generation and three chelate ligands (two with bridge length 4, one with bridge length 5) as well as one non-aromatic five-membered ring in the third generation and subsequently one aromatic six-membered ring in the fourth generation.

As already pointed out, the generation of new structures suffers from the inherent combinatorial explosion. While our approach allows for the full flexibility in discovering all possible structural features, it is crucial to build only structures which have a high chance of forming a suitable chemical environment for the reactive fragment. Therefore, we start...
with idealized ‘coordination geometries’ for every atom when adding new atoms for subsequent generations or introducing structural features such as ring systems or chelate ligands. We illustrate this principle at the example of a carbon atom. Here, a tetrahedral, trigonal planar, or linear ‘coordination geometry’ could be constructed. All these possibilities must be considered. The corresponding structures are then set up by applying standard bond lengths and angles [210]. It is sufficient to select some reasonable values for these structural parameters as the final starting structure will be subjected to a constrained optimization.

In general, MSD will try to cut as many structure branches of a structure tree as possible by invoking a split-and-prune principle (see the next subsections for examples). For this, the molecular space is partitioned into branches (‘split’) and each of the branches is then trimmed according to some rules (‘prune’).

### 3.4.5 Pruning of the Structure Tree

If a hydrogen atom, to be replaced by another ligand atom in the next generation, is connected to a base atom which has an additional hydrogen atom as direct neighbor (e.g., a methyl group) and if this atom group can be freely rotated, the hydrogen atoms will be considered equivalent. Consequently, redundant child structures which can be transformed into each other by rotating such groups are represented by only one child structure.

Particularly in later generations, two structures which were distinct in the generation process could turn out to be chemically equivalent after the constrained optimization. In order to avoid redundant structure branches, MSD identifies such redundant structures and merges the corresponding two structure branches into one distinct branch for the subsequent design step. Also in later generations, additional scaffold parameters such as the bite angle for chelate ligands can be specified.

Chemical wisdom may be applied to reduce the number of structures produced. Not all combinations of atom sets need to be allowed. For example, structural features such as a chain of four or more oxygen atoms or five and more nitrogen atoms may be ruled out in the generation process. Exceptions may be explicitly defined. If one is interested, for instance, in high-energy compounds, substructures with four or five
nitrogen atoms may be explicitly considered. Such rules can be switched on or off on input.

A structure may dissociate in the course of the constrained optimization. We interpret such a structure as thermodynamically unstable and hence as inappropriate to become a parent structure for subsequent generations. MSD then eliminates such a structure from the design process.

3.4.6 Graphical Explanation of the MSD

Fig. 3.9 illustrates the step-wise design process which is fully automated by the MSD. First, the structure of the activated fragment to be stabilized is defined. Although here we choose the structure of Ru-CO$_2$ from the full complex in Fig. 3.6, we note that any activation of a small molecule by structural distortion (change of bond lengths or bond angles) can produce an activated fragment. Moreover, the initial coordination geometry of the binding sites can be defined (e.g., octahedral, trigonal bipyramidal, or tetrahedral). Also on input, the scoring function(s) $\sigma_i$ as well as the corresponding threshold(s) $\lambda_i$ are set. The thresholds $\lambda_i$ can either be set as global parameters for the complete design process (as in this work) or chosen to be specific for each generation of structures. Here, we define as scoring functions the intrinsic coordination energy $D^\text{int}_e$ (obtained for structurally unrelaxed metal fragments and bent CO$_2$ ligand) and the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$.

The design process then starts with structure $s_{j=0}=S_0$ consisting of the chemical fragment and the binding sites which is fed into the first iteration of the MSD. Depending on the type of atoms and extension parameters defined in a configuration file, the MSD determines how many structures can be built by exploiting split-and-prune principles. The result of the split-and-prune design step is a set $\mathcal{G}$ of structures $s_{j \in \mathcal{G}}$, which is then subjected to a constrained optimization by means of standard quantum chemical software (in our case TURBOMOLE; see below) that preserves the structure of the activated fragment. For each structure $s_j$, the scoring functions are evaluated and only if the value for $\sigma_i$ of structure $s_j$ lies below the predefined threshold $\lambda_i$, the structure $s_j$ will be fed into a structure database.

Besides saving each structure that fulfills the thresholds together with its structure hierarchy, the structure database also identifies redundant structures as well as common structure branches and eliminates duplicate
**Figure 3.9:** Flowchart illustrating the individual steps of the MSD algorithm.
structures whenever necessary. The MSD then selects the next structure $s_{j'}$ out of the database to continue the subsequent iteration of the design process. Different selection criteria can be applied, such as choosing a structure $s_{j'}$ with as many multidentate ligands as possible or choosing a structure $s_{j'}$ without any carbon atoms in the second shell. For the study in this work, we simply select the structure with the lowest fragment gradient $|\nabla_{\text{frag}}E_{\text{el}}|$ for the next generation construction step without any additional constraints on the molecular topology.

### 3.5 Constructing the Target Complex

We now describe the shell-wise construction of the target complex as shown in Fig. 3.6. All structure branches leading to this target complex are shown in Fig. 3.10 with the corresponding numerical values for the fragment gradient in Table 3.1. As base structure, we chose a Ru($\eta^2$-CO$_2$) fragment with five binding sites in (idealized) octahedral symmetry.

Focusing on the leftmost branch in Fig. 3.10, our automated shell-wise construction algorithm places two phosphine ligands, a hydrogen atom, an ammine ligand, and an amide ligand at the binding sites around the ruthenium atom to form the structure [RuH(NH$_2$)($\eta^2$-CO$_2$)(PH$_3$)$_2$(NH$_3$)] which we label as $[1]_{\text{CO}_2}^5$. The fragment gradient $|\nabla_{\text{frag}}E_{\text{el}}|$ evaluates to 8.28·10$^{-2}$ hartree/bohr with the largest contribution coming from the $\eta$-bound oxygen atom of the CO$_2$ molecule with 2.96·10$^{-2}$ hartree/bohr. The intrinsic binding energy $D_{\text{e}}^{\text{int}}$ of this structure is -194.5 kJ/mol. Note that the distortion energy required to prepare a free CO$_2$ molecule in the bent structure ready for coordination is calculated to be 140.3 kJ/mol, which significantly reduces the intrinsic binding energy.

In general, the MSD also constructs some structures with positive intrinsic binding energy for the CO$_2$ fragment in the first generation which we do not consider for subsequent construction steps. From the second generation onwards, all generated structures exhibit a negative intrinsic binding energy such that the overall design task solely focuses on reducing the fragment gradient.

In the next step, the MSD methylates both phosphorus atoms as well as the amide ligand symmetrically and extends the ammine ligand by one methyl group and one amide ligand, respectively. The resulting structure [RuH(N(CH$_3$)$_2$)($\eta^2$-CO$_2$)(P(CH$_3$)$_3$)$_2$(NH(CH$_3$)NH$_2$)] we label as $[1 \, 1]_{\text{CO}_2}^5$.
Figure 3.10.: Overview of all structure branches which lead to the target structure at the bottom.
Table 3.1: Absolute values of the Cartesian gradients in hartree/bohr on the atoms of the activated fragment (O(2) is the oxygen atom \(\eta\)-bound to the ruthenium atom) for the different molecular environments shown in Fig. 3.10. For all molecular environments, the distances Ru-C, C-O₁, C-O₂ and the angle O-C-O were fixed to the values given Section A.1.2. Structures marked with an asterisk were calculated with geometry constraints in Cartesian coordinates, where only the ruthenium atom and the CO₂ moiety were positionally constrained. The intrinsic electronic coordination energy \(D_e^{\text{int}}\) is given in kJ/mol.

<table>
<thead>
<tr>
<th>ID</th>
<th>(\nabla_{\text{R}<em>1}E</em>{\text{el}})</th>
<th>(\nabla_{\text{C}<em>1}E</em>{\text{el}})</th>
<th>(\nabla_{\text{O}<em>1}E</em>{\text{el}})</th>
<th>(\nabla_{\text{O}(2)}E_{\text{el}})</th>
<th>(\nabla_{\text{frag}}E_{\text{el}})</th>
<th>(D_e^{\text{int}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>([1]^{5}\text{CO}_3) *</td>
<td>1.66 \times 10^{-2}</td>
<td>2.00 \times 10^{-2}</td>
<td>1.67 \times 10^{-2}</td>
<td>2.96 \times 10^{-2}</td>
<td>8.28 \times 10^{-2}</td>
<td>-194.5</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2) *</td>
<td>1.68 \times 10^{-2}</td>
<td>9.45 \times 10^{-3}</td>
<td>2.39 \times 10^{-3}</td>
<td>2.01 \times 10^{-2}</td>
<td>4.88 \times 10^{-2}</td>
<td>-234.8</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>4.68 \times 10^{-3}</td>
<td>6.11 \times 10^{-3}</td>
<td>3.95 \times 10^{-3}</td>
<td>5.07 \times 10^{-3}</td>
<td>1.98 \times 10^{-2}</td>
<td>-262.4</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>4.89 \times 10^{-3}</td>
<td>6.50 \times 10^{-3}</td>
<td>5.47 \times 10^{-3}</td>
<td>5.34 \times 10^{-3}</td>
<td>2.22 \times 10^{-2}</td>
<td>-262.6</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>6.05 \times 10^{-3}</td>
<td>6.66 \times 10^{-3}</td>
<td>4.80 \times 10^{-3}</td>
<td>5.65 \times 10^{-3}</td>
<td>2.32 \times 10^{-2}</td>
<td>-257.9</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>5.45 \times 10^{-3}</td>
<td>6.22 \times 10^{-3}</td>
<td>5.98 \times 10^{-3}</td>
<td>5.20 \times 10^{-3}</td>
<td>2.28 \times 10^{-2}</td>
<td>-256.4</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>5.66 \times 10^{-3}</td>
<td>6.48 \times 10^{-3}</td>
<td>4.78 \times 10^{-3}</td>
<td>5.50 \times 10^{-3}</td>
<td>2.24 \times 10^{-2}</td>
<td>-261.9</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>6.14 \times 10^{-3}</td>
<td>3.98 \times 10^{-3}</td>
<td>1.15 \times 10^{-3}</td>
<td>2.22 \times 10^{-3}</td>
<td>1.35 \times 10^{-2}</td>
<td>-243.8</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>5.99 \times 10^{-3}</td>
<td>3.86 \times 10^{-3}</td>
<td>1.21 \times 10^{-3}</td>
<td>2.68 \times 10^{-3}</td>
<td>1.38 \times 10^{-2}</td>
<td>-242.2</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>6.27 \times 10^{-3}</td>
<td>4.52 \times 10^{-3}</td>
<td>1.01 \times 10^{-3}</td>
<td>3.07 \times 10^{-3}</td>
<td>1.49 \times 10^{-2}</td>
<td>-240.8</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>6.77 \times 10^{-3}</td>
<td>4.28 \times 10^{-3}</td>
<td>1.24 \times 10^{-3}</td>
<td>2.83 \times 10^{-3}</td>
<td>1.51 \times 10^{-2}</td>
<td>-242.2</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>6.10 \times 10^{-3}</td>
<td>4.27 \times 10^{-3}</td>
<td>1.20 \times 10^{-3}</td>
<td>4.01 \times 10^{-3}</td>
<td>1.56 \times 10^{-2}</td>
<td>-237.4</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>6.34 \times 10^{-3}</td>
<td>4.43 \times 10^{-3}</td>
<td>1.04 \times 10^{-3}</td>
<td>3.05 \times 10^{-3}</td>
<td>1.49 \times 10^{-2}</td>
<td>-240.6</td>
</tr>
<tr>
<td>([1 1]^{5}\text{CO}_2)</td>
<td>2.25 \times 10^{-3}</td>
<td>2.67 \times 10^{-3}</td>
<td>1.41 \times 10^{-3}</td>
<td>5.65 \times 10^{-4}</td>
<td>6.89 \times 10^{-3}</td>
<td>-258.0</td>
</tr>
</tbody>
</table>
in Fig. 3.10. The fragment gradient decreases to $4.88 \cdot 10^{-2}$ hartree/bohr compared to its parent structure. Already at this early stage of our shell-wise construction, the formation of chelate ligands is possible. The MSD framework forms a four-atomic bidentate ligand out of the two NR-CH$_3$ arms to yield the structure indexed by $[1\ 1c_4^4\ 1a_1^5]_{CO_2}$. For the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$, we calculate $1.98 \cdot 10^{-2}$ hartree/bohr which is considerably smaller than $|\nabla_{\text{frag}} E_{\text{el}}|$ of the parent structure.

The MSD now replaces one hydrogen atom of each methyl group of the chelate ligand by a CH$_3$ group and adds another CH$_3$ group to the methyl group directly adjacent to the first generation hydrogen atom to form $[1\ 1c_4^4\ 1a_1^5]_{CO_2}$. Compared to its parent structure $[1\ 1c_4^4]_{CO_2}$, the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ increases to $2.22 \cdot 10^{-2}$ hartree/bohr.

At this stage of the molecular scaffold construction, the creation of saturated and partially saturated five-membered ring topologies is possible. One structure forms a pyrazole-like ring topology adjacent to the sole hydrogen atom represented by $[1\ 1c_4^4\ 1a_1^5]_{CO_2}$. As a result of having a five-membered aromatic ring in the molecular scaffold, the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ decreases to $1.35 \cdot 10^{-2}$ hartree/bohr, hence reducing it by almost a factor of two. In contrast to all previous parent structures of the structure branch followed, the ruthenium atom contributes most to the overall fragment gradient with $6.14 \cdot 10^{-3}$ hartree/bohr.

Replacing one hydrogen atom of the terminal methyl group of the NR-CH$_2$CH$_3$ arm yields the structure $[1\ 1c_4^4\ 1a_1^5\ 1]_{CO_2}$ in the fourth generation. As this methylation occurs at the largest possible distance to the Ru($\eta^2$-CO$_2$) fragment, we find the fragment gradient to be $1.49 \cdot 10^{-2}$ hartree/bohr, which is slightly larger than the one of its parent structure.

Starting with the fourth generation, the formation of aromatic as well as non-aromatic six-membered ring topologies is possible. The MSD framework constructs a pyridine-like ligand out of the two remaining molecular arms which, together with the pyrazole-like ring, forms a large bidentate pyridinyiazolato ligand identified in $[1\ 1c_4^4\ 1a_1^5\ 1a_1^5]^5_{CO_2}$. This molecular scaffold is the target complex given in Fig. 3.6 with a fragment gradient of $|\nabla_{\text{frag}} E_{\text{el}}| = 6.89 \cdot 10^{-3}$ hartree/bohr. As in the case of its parent structure, the ruthenium atom contributes most to the overall fragment gradient with $2.25 \cdot 10^{-3}$ hartree/bohr. This again shows the beneficial effect of a ring topology on the overall fragment gradient compared to alkyl substitutions in previous generations. Note, however, that the fragment is not closer to $1 \cdot 10^{-3}$ because the activated fragment is still structurally
Constrained and its structure does not represent a minimum structure for
the DFT settings chosen in this study. The intrinsic binding energy \( D_{\text{e}int} \)
of the target complex is -258.30 kJ/mol and therefore very similar to those
calculated for all other structures in the same branch (see Table 3.1).

### 3.5.1 Position of the Target Branch in the Screened Molecular Space

As described above, already a few contiguous additions of atoms together
with chelate and ring topologies allow for the design of diverse molecular scaffolds. This diversity is depicted in Fig. 3.11 where we show the relation between the fragment gradient \( |\nabla_{\text{frag}} E_{\text{el}}| \) and the intrinsic binding energy \( D_{\text{e}int} \) for the first four generations of 45’486 shell-wise constructed complexes. The candidate structures in the first generation are characterized by a wide spread of the fragment gradient (up to 0.4 hartree/bohr) and of the intrinsic binding energy (from -400 kJ/mol to +100 kJ/mol). For the second and third generation, the range of possible fragment gradients is reduced to [0, 0.2] hartree/bohr and [0, 0.08] hartree/bohr, respectively. Particularly for the intrinsic binding energy, the MSD designs candidate structures with only negative intrinsic binding energies between -330 kJ/mol and -160 kJ/mol in the second generation and between -320 kJ/mol and -170 kJ/mol in the third generation.

Hence, the molecular scaffolds show increasingly vanishing effects with reduced spread in higher generations. In the fourth generation, the spread of \( |\nabla_{\text{frag}} E_{\text{el}}| \) decreases further to 0.04 hartree/bohr and all generated structures show a \( D_{\text{e}int} \) between -270 kJ/mol and -190 kJ/mol.

While it is hard to classify all generated structures in previous generations, this is now possible in the fourth generation. The larger set of structures can be categorized as molecular scaffolds with and without hydrogen bonds to the CO\(_2\) ligand. Those without hydrogen bonds are represented by the standard deviational ellipse [211] in green in Fig. 3.11. A smaller subset of structures shows one or more hydrogen bonds between atoms of the molecular scaffold and the CO\(_2\) ligand; these structures are represented by the standard deviational ellipse in gray.
Figure 3.11.: Fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ in relation to the intrinsic binding energy $D_{\text{el}}^{\text{int}}$ for the first four generations. Two distinct subsets (with and without hydrogen bonds to the CO$_2$ moiety) of structures can be identified in the fourth generation, represented by the standard deviational ellipses in gray and green, respectively.

Figure 3.12.: Left: Binding sites of the target complex for the functionalization step. Right: The set of functional groups adopted, from top to bottom: CO$_2$, CH-CH$_2$, COOH, pentyl, phenyl, and pyrol. The atoms directly connected to the molecular scaffold are marked in green.
3.5.2 Functionalization as a Means to Further Lower the Fragment Gradient

Once a molecular scaffold is built around the activated fragment and the overall fragment gradient is below a certain threshold, the MSD functionalizes the now existing structures by pre-optimized molecular fragments instead of continuing the atom-wise build-up of the molecular environment to further reduce the fragment gradient $|\nabla_{\text{frag}} E_{el}|$. To this end, we generate all possible structures starting from the structure $[1 \, 1c \, 1a_1^5 \, 1a_1^6]_\text{CO}_2$ by permutationally replacing all hydrogen atoms in generation four and above by a selected set of functional groups. The hydrogen atoms to be replaced and the functional groups employed for the functionalization step are shown in Fig. 3.12.

The resulting distribution of fragment gradients and intrinsic binding energies for all emerging structures is shown in Fig. 3.13. As expected, the fragment gradient cannot be significantly lowered for scaffold modifications that are at a distance to the activated fragment. The average values of $|\nabla_{\text{frag}} E_{el}|$ and $D_e^{\text{int}}$ for all generated structures are $1.25 \cdot 10^{-2}$ hartree/bohr and $-251.3$ kJ/mol, respectively, and therefore slightly higher than the reference values of the constrained-optimized target complex ($6.89 \cdot 10^{-3}$ hartree/bohr and $-258.0$ kJ/mol).

The four structures with the lowest fragment gradient in this generation are shown in Fig. 3.14. Further functionalization on hydrogen atoms in lower generations could possibly lead to lower fragment gradients. When comparing the progression of fragment gradients over all five generations, the spread decreases significantly as can be seen on the right hand side of Fig. 3.14.

3.6 Designing Alternative Candidate Structures

So far, the analysis focused on the structure branch which leads directly to our target complex. Now, we take advantage of the flexibility of the MSD framework and design other candidate structures. To this end, we choose the structure with the lowest fragment gradient in each generation as parent structure for the following design step while we keep the same atom set (H, C, N, O, P) as before for replacing the binding sites in each
Chapter 3  |  MOLECULAR SCAFFOLD DESIGN BY GDMC

Figure 3.13.: Left: Fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ in relation to the intrinsic binding energy $D^\text{int}_e$ for the fifth generation, i.e., the functionalization step. The standard deviational ellipse is shown in green. Right: Progression of $|\nabla_{\text{frag}} E_{\text{el}}|$ over the first five generations of molecular scaffolds.

Figure 3.14.: Overview of the best four functionalized molecular scaffolds employing the target complex as base structure. Both Lewis structures (with $R=P(\text{CH}_3)_3$) and constrained-optimized structures are presented. In parentheses, values for the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ and the binding energy are given in hartree/bohr and kJ/mol, respectively, for both the constrained (C) optimized structures ($D^\text{int}_e$) and the fully (F) optimized structures ($D_e$).
Designing Alternative Candidate Structures

3.7

generation. Figure 3.15 shows this structure branch and the corresponding candidate structures.

In the first generation, the ligand environment with two NH₃ ligands, a hydrogen atom, a methyl group, and a phosphine ligand with their relative positions as indicated in Fig. 3.15 has the lowest fragment gradient with a value of $|\nabla_{\text{frag}E_{\text{el}}}| = 3.78 \cdot 10^{-3}$ hartree/bohr. As the value for the binding energy with $D_{\text{e}}^{\text{int}} = -267.8$ kJ/mol is also very good, we continue the design process with this molecular scaffold as parent structure. From all the child structures, the structure $[5\ 1]^5_{\text{CO}_2}$ has the lowest fragment gradient with $|\nabla_{\text{frag}E_{\text{el}}}| = 1.56 \cdot 10^{-3}$ hartree/bohr.

From our first design step in Section 3.5, we understand the necessity for chelate and cyclic topologies in the molecular scaffold. The MSD therefore selects the structure $[5\ 1]^5_{\text{CO}_2}$, which is the structure with the lowest fragment gradient in the second generation, and builds all possible structures containing chelate or cyclic topologies before adding additional atoms. Out of these resulting structures, the structure $[5\ 1c_1^4]^5_{\text{CO}_2}$ has the smallest fragment gradient with $|\nabla_{\text{frag}E_{\text{el}}}| = 3.37 \cdot 10^{-3}$ hartree/bohr. Even though its fragment gradient is almost twice as large as the one of structure $[5\ 1]^5_{\text{CO}_2}$, the MSD selects this structure as parent structure for the next iteration. For feasibility reasons, we changed our criterion for selecting an appropriate structure to continue the design process at this stage as follows: We select the structure with the lowest fragment gradient out of all structures in the second generation with at least one chelate or cyclic topology. This heuristic rule forces the MSD to follow a structure branch containing chelate or cyclic topologies instead of following other structure branches with chain-like ligands and possibly lower fragment gradients. The MSD then identifies the child structure $[5\ 1c_1^4\ 1]^5_{\text{CO}_2}$ as the one with the lowest fragment gradient of $|\nabla_{\text{frag}E_{\text{el}}}| = 1.89 \cdot 10^{-3}$ hartree/bohr.

Note, however, that not all structures of the next generation have been explicitly considered in the screening process. Depending on the number of generations that might still need to be considered, a huge number of additional structures can be generated, which we do not consider in this work. In order to reduce the effort at this stage, one might want to restrict the atom set to just (C, H) in higher generations (as otherwise several hundred thousand structures need to be inspected) or probe structures in further iterations by an evolutionary algorithm [212].
Figure 3.15: Structure branch for another candidate structure which stabilizes the CO$_2$ moiety. In parentheses, values for the fragment gradient $|\nabla_{\text{frag}}E_{\text{el}}|$ and the binding energy are given in hartree/bohr and kJ/mol, respectively, for both the constrained (C) optimized structures ($D_{\text{e}c}^{\text{in}}$) and the fully (F) optimized structures ($D_{\text{e}}$). The low coordination energy $D_{\text{e}}$ for structure $[5 1]_{\text{CO}_2}^5$ arises due to the phosphane group being closer to the ruthenium atom after optimization of the ligand environment without the CO$_2$ fragment (243.5 pm instead of 326.7 pm).
3.7 Comparison with Similar Ligand Environments

In the target complex as shown in Fig. 3.6, the unsaturated pyridinylazolato ligand is a peculiar structural component. In Fig. 3.16, we show eight alternative ligand environments which are similar to the target complex. The corresponding fragment gradients $|\nabla_{\text{frag}} E_{\text{cl}}|$ and intrinsic coordination energies $D_{\text{e}}^{\text{int}}$ are collected in Table 3.2. Compared to these eight alternative ligand environments, the target complex $[1 1c_1^1 1a_1^5 1a_1^6]_{CO_2}$ has the smallest fragment gradient. With a fragment gradient of $9.21 \times 10^{-3}$ hartree/bohr, the structure $[1 2c_1^4 1 1a_1^6]_{CO_2}$ with a bispyridyl ligand ranks second. The structure $[1 3c_1^4 1a_1^5 1a_1^6]_{CO_2}$ has the same pyridinylazolato ligand as the target complex, however, in an inverted form. With a fragment gradient of $9.55 \times 10^{-3}$ hartree/bohr, this structural arrangement is less favorable than the one of the target complex. We note that nitrogen appears to be an important ligand atom to be bound directly to the transition metal center as the fragment gradient for all structures containing at least one nitrogen atom is smaller than for those three structures containing no nitrogen atom in their ligand environment (i.e., with a fluorene, a biphenyl, and a phenyl-pentyl ligand environment). In particular, the nitrogen atom opposite to the sole hydrogen atom in the first ligand sphere appears to have a decreasing effect on the fragment gradient. This is evident by comparing structure $[2 1c_1^4 1a_1^5 1a_1^6]_{CO_2}$ with structure $[2 1c_1^4 1 1a_2^6]_{CO_2}$. Both structures contain a phenyl-pyridinyl ligand, however, with a different stereogeometry. Here, the structure $[2 1c_1^4 1a_1^5 1a_1^6]_{CO_2}$ with the pyridinyl ligand opposite to the sole hydrogen atom has a fragment gradient of $1.71 \times 10^{-2}$ hartree/bohr which is lower than the fragment gradient of structure $[2 1c_1^4 1 1a_2^6]_{CO_2}$, which is $2.51 \times 10^{-2}$ hartree/bohr.
Table 3.2: Absolute values of the Cartesian gradients in Hartree/bohr on the atoms of the activated fragment (O(2) is the oxygen atom of the CO fragment). Where the sole hydrogen atom rotates around the ruthenium atom after optimization of the ligand environment without ruthenium atom before optimization of the ligand environment without the CO fragment. The same trend is visible for the structure given in Ref. [5]. The low coordination energy $D^* \Delta$ for structure [1] due to the phenyl group resulting around the Ru-C-C-O bond to the ruthenium atom for the different molecular environments shown in Fig. 3.10. For all molecular environments, the distances $r$-bound to the ruthenium atom of the actives and Ref. [5].

| $D^*$ | $\Delta$ | $r$-bound $\Delta$ | $|F_\text{Hartree/bohr}|$ | $F_\text{Hartree/bohr} \Delta$ | $r$-bound $\Delta$ | $|F_\text{Hartree/bohr}|$ | $F_\text{Hartree/bohr} \Delta$ | $r$-bound $\Delta$ |
|-------|---------|--------------------|----------------------------|----------------------------|--------------------|----------------------------|----------------------------|--------------------|
| 3.4  | 1.5     | 0.4                | 0.3                        | 0.3                        | 0.4                | 0.3                        | 0.3                        | 0.4                |
| 3.2  | 1.4     | 0.3                | 0.2                        | 0.2                        | 0.3                | 0.2                        | 0.2                        | 0.3                |
| 3.0  | 1.3     | 0.2                | 0.1                        | 0.1                        | 0.2                | 0.1                        | 0.1                        | 0.2                |
| 2.8  | 1.2     | 0.1                | 0.0                        | 0.0                        | 0.1                | 0.0                        | 0.0                        | 0.1                |

| $D^*$ | $\Delta$ | $r$-bound $\Delta$ | $|F_\text{Hartree/bohr}|$ | $F_\text{Hartree/bohr} \Delta$ | $r$-bound $\Delta$ | $|F_\text{Hartree/bohr}|$ | $F_\text{Hartree/bohr} \Delta$ | $r$-bound $\Delta$ |
|-------|---------|--------------------|----------------------------|----------------------------|--------------------|----------------------------|----------------------------|--------------------|
| 3.4  | 1.5     | 0.4                | 0.3                        | 0.3                        | 0.4                | 0.3                        | 0.3                        | 0.4                |
| 3.2  | 1.4     | 0.3                | 0.2                        | 0.2                        | 0.3                | 0.2                        | 0.2                        | 0.3                |
| 3.0  | 1.3     | 0.2                | 0.1                        | 0.1                        | 0.2                | 0.1                        | 0.1                        | 0.2                |
| 2.8  | 1.2     | 0.1                | 0.0                        | 0.0                        | 0.1                | 0.0                        | 0.0                        | 0.1                |

The CO$^2$ fragment where the sole hydrogen atom rotates around the ruthenium atom after optimization of the ligand environment without ruthenium atom before optimization of the ligand environment without the CO$^2$ fragment. The same trend is visible for the structure given in Ref. [5]. The low coordination energy $D^* \Delta$ for structure [1] due to the phenyl group resulting around the Ru-C-C-O bond to the ruthenium atom for the different molecular environments shown in Fig. 3.10. For all molecular environments, the distances $r$-bound to the ruthenium atom of the actives and Ref. [5].
Comparison with Similar Ligand Environments

Figure 3.16.: Ligand environments similar to the target complex as shown in Fig. 3.6. The structures are ordered from left to right by increasing fragment gradient $|\nabla_{\text{frag}} E_{el}|$. 
4
Conditional Design of Molecular Scaffolds

Until now, we have designed single structures in order to achieve one specific design goal which was the stabilization of an (activated) chemical fragment. As a next step, we consider the highly complicated case of molecular function expressed as a network of molecular-rearrangement steps, in which a large number of stable intermediates rather than a single structure is to be considered. In such a case, the rationalization and understanding of function is achieved in terms of a network of molecular structures. Designing function then implies designing components (nodes) of the network. Such components can be stable intermediates and transition states. On Born–Oppenheimer electronic-energy surfaces, they represent stationary points and can be considered as nodes of the network. In a design process, structure types and decorated (derivatized) scaffolds are to be identified that fulfill specific roles at different nodes of the network.

An example for such a network that fulfills a chemical function is a catalytic cycle. Given this function, the generic form of a catalytic cycle can be anticipated, even if a structural realization of the catalyst is not known and to be found. The catalyst transforms one molecular species into another one and possible intermediates can be anticipated in a prototypical catalytic cycle, while the detailed structure of the catalyst represents the ultimate goal of a design process. The fact that a catalytic net reaction may be accomplished through different mechanisms does not counteract
the idea of a generic transformation network as all potential mechanisms can be anticipated by applying conventional chemical concepts.

The prototypical cycle might be simplified by excluding transitions through nodes that are known to be always feasible in a given context (e.g., reaction steps that are known to always have a negative free reaction energy under pre-defined reaction conditions). Then, the prototypical cycle stripped by such nodes that are not key to the chemical process represents the generic network that we consider for our design approach. Hence, the network of nodes that we deem essential for achieving some design goal is not to be confused with a chemical reaction network that maps all elementary reaction steps of a detailed reaction mechanism.

Consequently, the edges that may connect nodes in such a generic network shall only be understood as a means to establish a certain contextual dependence of the nodes. A reaction flow may pass nodes according to this ordering defined by the connecting edges. Edges can be important for the design process as structural changes from node to node might be large so that structural constraints imposed on different nodes in a design process must not be mutually exclusive. If a structural transgression from node to node coded by the edges introduces structural changes, competing structural requirements are possible.

In general, many nodes in a network will be required to map out the essential functionality of a molecular process. Even more severe, the predictive design of a chemical process might also require to consider all possible side reactions that could deteriorate the function of the molecular design under certain reaction conditions. Such side reactions then need to be made unaccessible by structural design. As a consequence, a design process is, in general, very complicated, not only because the theoretical modelling might require significant computational resources, but because of the sheer number of structures that need to be considered for the nodes of a generic network.

Clearly, designing individual nodes separately may compromise design achievements that have already been made for other nodes. In other words, the design of molecular scaffolds that support the chemical function represented by the network is a highly entangled process that aims to fulfill all constraints known for the individual nodes by the same type of molecular structure. Here, we specifically address the problem that the optimization of a molecular scaffold to fulfill properties of a single node
in a network cannot be carried out independently of the optimization of this type of molecular scaffold for all other nodes in the same network.

As an example for this, Elward and Rinderspacher recently developed a deterministic optimization procedure which is able to handle multiple constraints [213] which they applied to a rigid aromatic system. They then generated new structural alternatives by permutationally adding four different substituents to eight predefined substitution sites. In our case, our conditional optimization problem is much more complex since we start the generation of new structures from scratch, while allowing all chemically imaginable structural features during the generation process.

While this fact that a molecular scaffold must fulfill different chemical properties at different nodes can be considered a substantial drawback for rational compound design, we propose that an optimization procedure should exploit these conditions in order to eventually accelerate the design process. We call this problem the **conditional design** problem.

Our aim is the development and implementation of a general strategy for conditional design. While we appreciate the high complexity of this endeavor — especially considering a subsequent experimental realization of a theoretically designed scaffold —, we consider our approach a valuable tool to generate ideas for new molecular scaffolds and to identify structure–function relationships.

### 4.1 Principles of Conditional Design

We start from a generic network of nodes whose structural representation is unknown and shall be found in the design process. As the network is an abstract representation of the chemical function, it is *not* to be confused with a detailed reaction network of elementary steps that map the stationary points of a potential energy surface. As such, the generic network needs to comprise only the most critical steps of a chemical process. Steps that are known to be accomplishable by any molecular structure designed need not be part of the generic network.

Hence, the network may be formulated in a very general way so that no explicit molecular structures are assigned to the nodes. Instead a general notion is required of what the node is expected to perform in order to make a generic chemical process viable. The chemical functions that can be realized by the network will be encoded by energies and molecular
properties calculated for a molecular structure assigned to a node. Ideally, these energies should be free energies, which, however, are not easy to calculate under all circumstances. If all (relevant) transformation steps are governed by the change in electronic energy rather than by entropic or nuclear-motion effects, the Born–Oppenheimer surface may be taken as a reliable substitute for the free-energy hypersurface. Then, the generic network represents those chemical transformation steps encoded by sequences of stationary points that are decisive for the chemical process under consideration. Yet, the approximate nature of electronic structure methods as well as kinetic models requires the introduction of free energy uncertainties in order to obtain a quantitative understanding of such complex chemical reaction networks. [7].

We now aim to design a molecular scaffold structure that is able to feature a certain property for one specific node of the network under the constraint that it also fulfills constraints imposed at the other nodes in the network, which we denote conditional design:

**Conditional Design:** In a generic network of nodes that map critical steps of some general molecular transformation process, the molecular structure to accomplish the function of one node should be constructed in such a way that structural constraints required at all other nodes are obeyed on-the-fly during the design process.

In the following subsections, we first consider elements of conditional design at the example of a toy network. In a later step, we apply conditional design to the Schrock dinitrogen activating molybdenum complex. For the sake of simplicity, we restrict ourselves to the electronic energy as a target quantity, i.e., to processes which can be well described on a Born–Oppenheimer potential energy surface. Our design principle of choice within the conditional design framework is again the GdMC concept [77]. As explained in chapter 3, GdMC aims at minimizing the structure gradient of the electronic energy on all nuclei of the full target molecule, however, in case of conditional design it can be replaced by any other design principle.
**Figure 4.1.:** Top: A generic network: Nodes corresponding to minima (‘MIN’) are highlighted by blue circles, while nodes that represent transition states (‘TS’) are printed in green circles. The black line between two nodes indicates an edge that provides ordering of nodes in time-wise sequential processes. **Bottom:** A small reaction network of two minimum structures connected by a transition state.

### 4.1.1 The Generic Network

Structural conversions of a molecular system are associated with an energy change expressed as moving on an energy hypersurface. Decisive structures for such structural rearrangement processes can be assigned according to various theoretical concepts of which stationary-point criteria [214, 215] and Markov state models [216–218] are two examples. Since dynamical processes connect all such structures, they can be arranged as nodes in a network. A pictorial representation of such a network is given in Fig. 4.1. The minima (‘MIN’) can be reached through a transition state (‘TS’).

We consider the simplest such network, in which only two nodes are connected by a transition state (see bottom part of Fig. 4.1). The goal of conditional design is now the construction of a molecular environment facilitating the given chemical process represented by this toy network with certain properties assigned to all nodes.
Figure 4.2.: Sketch of how the Schrock-type dinitrogen fixation complex can be built up in a shell-wise fashion.

4.1.2 Shell-wise Construction of Molecular Scaffolds

Within the conditional design concept, one specific rational compound design approach is chosen, but conditional design is not restricted to a specific choice of method. It is the mathematical formalism of the design approach chosen in which the desired property is encoded. Depending on the design method, molecular scaffolds can be built from chemically meaningful building blocks (such as chelate ligands or functional groups) or sequentially shell by shell. The molecular environment may even be realized by reconstructing a continuous environment potential that eventually needs to be represented by a stable chemical structure in a two-step optimization [76].

We define the smallest chemical fragment representing the desired property at one node. Such a procedure is meaningful if local chemical processes dominate the chemical function of the network. For example, in Fig. 4.2, the desired property is represented by a metal binding the inert molecule dinitrogen, together forming the chemical fragment. The function 'binding dinitrogen' is performed by only a rather small number of known transition metal complexes and therefore it represents a design target formulated as a node in some network that considers transformation of inert dinitrogen. The design process now needs to identify a suitable metal atom in a suitable ligand environment to accomplish binding of dinitrogen. The latter, i.e., the construction of a ligand environ-
ment represents a formidable task. Its construction may be conveniently divided into two phases: binding new atoms to produce different (chelate) scaffolds of increasing size and complexity and then derivatizing them by introducing molecular fragments (functional groups).

As shown in Fig. 4.2, the local reactivity center is first surrounded by atoms belonging to the first shell. Then, the next shell of atoms is added, shown in orange and green in Fig. 4.2. This procedure is carried out by adding more and more shells of atoms until the desired property is fulfilled by the constructed scaffold, i.e., we apply the MSD framework introduced in chapter 3.

4.1.3 Structure Trees and the Split and Prune Method

We emphasize again that the generation-wise build-up of a molecular environment implies a hierarchy with a base structure and the possibility to trace the parent structure of each child structure all the way back to the base structure. In the process of building up such a structure hierarchy, it is important to abide by one principle, namely that one specific structure is created once and only once. Only then is it possible to dissect the structure hierarchy into distinct branches.

The aim of conditional design will then be the elimination of certain structure branches by cutting them as close as possible to the tree trunk ('prune') in order to make the whole design process structurally extensive, but still computationally feasible. The criteria for such an elimination will be set by the other nodes in the network. Clearly, the elimination of branches must be feasible before the whole branch is explicitly evaluated in order to be efficient. The 'split and prune' procedure (see chapter 3.4.5) will allow us to significantly reduce the computational effort for scaffold design. The specific design problem under consideration determines how many branches and how close to the trunk the structure tree can be thinned out.
4.1.4 The Property Classes

In the course of the design process, more than one structure satisfying the pre-defined property may be found. To collect all such structures we introduce the concept of a property class \( C_i \). As the design process will be carried out for each individual node of a network separately, there will be at least one property class for each node and all structures with that property at node \( i \) are member of the property class \( C_i \).

In practice, we start the design process at a selected node, for which the conditions of other nodes can be most easily implemented. For instance, the starting node could be minimum 1 of our toy reaction network (see Fig. 4.3). At this specific node, the structure to be found must exhibit a desired property which is directly encoded in the design principle chosen.

During the design process, certain structures will be found that feature the desired property, that is, they belong to the property class \( C_{MIN} \). The goal is now to ensure that only those structures are actually constructed and evaluated which feature the properties of the other nodes as well.

4.1.5 Conditional Design for a Generic Chemical-Process Network

Once the design process is finished for one node and structures featuring the desired property have been found, the focus will then be on all the other nodes of the network; in our toy case, we focus on the transition state and on minimum 2 in Fig. 4.3.

Collecting the set of all structures that exhibit the desired property corresponding to node \( i \) into the property class \( C_i \), we search for one or more structures \( s_i \) that belong to the class \( C_2 \) conditioned on the fact that they also belong to the class \( C_i \), that is,

\[
S = \{ s_i \in C_2 | s_i \in C_i \} \quad (4.1)
\]

This condition can be generalized to a reaction network with \( n \) nodes (such as in Fig. 4.1) straightforwardly: Focusing on node \( n \) of a chemical reaction network, we look for structures that belong to class \( C_n \) conditioned on the fact that they also belong to the classes \( C_1, C_2, ..., C_{n-1} \):

\[
S = \{ s_i \in C_n | s_i \in (C_1 \cap C_2 \cap ... \cap C_{n-1}) \} \quad (4.2)
\]
For the transition state 1 ('TS\(_a\)') of our toy network, this means that we rationally design the structures featuring the desired property given at this node. Instead of designing all possible structures which are able to accomplish this, we only focus on structure trees belonging to \(\mathcal{C}_{\text{MIN}_1}\). By doing this, we can eliminate the remaining structure branches which could possibly be part of \(\mathcal{C}_{\text{TS}_1}\), yet do not belong to \(\mathcal{C}_{\text{MIN}_1}\). As shown in Fig. 4.3, the structure branches to be constructed for 'TS\(_1\)' are highlighted and correspond to the three structure branches that return a molecular structure belonging to \(\mathcal{C}_{\text{MIN}_1}\). All the other remaining structure branches constructed for 'MIN\(_1\)' and which do not yield a member structure of \(\mathcal{C}_{\text{MIN}_1}\) will not be considered any more. It should be noted that, in principle, more than one structure branch can start from each colored node in Fig. 4.3. However, for the sake of clarity, only one child structure is formed in each generation.

We now assume that out of the three remaining structure branches at 'TS\(_a\)', only two of them yield molecular representations featuring the desired property at 'TS\(_a\)' (shown by circles in light green at the end of each structure branch around 'TS\(_a\)'.). Consequently, if we continue the design
procedure at minimum 2 (‘MIN2’), we only need to reconstruct those two structure branches at this node of our reaction network. In our example in Fig. 4.3, only one structure branch at ‘MIN2’ yields a child structure belonging to $C_{\text{MIN2}}$ and thus returns a molecular structure featuring the desired property given for this specific node. We now have successfully rationalized a molecular representation of the chemical environment that is able to feature each specific property at any given node of our reaction network.

4.1.6 Expression of Conditions in Terms of Local Descriptors

Hence, the definition of descriptors that map constraints on all nodes but node $i$ allows us to reduce the number of potentially interesting structures in the design process at node $i$. The first node to choose for the design process is therefore a critical choice for the overall efficiency of the design process. It is to be selected based on how easy it is to encode constraints on all other nodes into a descriptor that can be evaluated at node $i$. These descriptors should be easy to evaluate, local, and transferable (from the other nodes of the network to node $i$ for which a scaffold is constructed).

We will discuss in detail at our specific example below that locality of quantum mechanical descriptors is most desirable because our shell-wise structural growth will otherwise be difficult to converge. Structural modifications at a spatial distance from the base structure could produce significant changes in a non-local descriptor rendering the truncation of tree branches impossible.

Moreover, it may be anticipated that too rigid constraints will narrow down the design options too much so that no flexibility can be presented for the other nodes. It is important to understand that the descriptors that can be used for the design process are neither strict nor exact. They can only be a guideline, a chemical concept that allows us to avoid the full calculation at the other nodes. Consequently, a rather wide error margin should be taken into account when thinning out the structure tree. This process is a delicate balance between computational feasibility and allowance of an error range for the fulfillment of a local quantum descriptor that is evaluated for a structure at node $i$, but that was extracted
for a yet to be explicitly calculated structure at some other node in the network.

Clearly, this procedure demands that local descriptors attached with a reasonable error margin can be defined prior to the design process. Once the design process at one node is finished, the selected structures may be presented as optional scaffolds for explicit quantum chemical calculation at the other nodes. Depending on their number, one may first select one of the other nodes to further narrow down the structures in explicit quantum chemical calculations or directly subject it to all other nodes in a trivially parallel manner. Since all such structures optimized for node \( i \) already fulfill the descriptors defined by the other nodes, further conditional design is not necessary and a two-phase process emerges for design: conditional design in the first phase for a most suitable node followed by structural evaluation at all other nodes in the second one.

4.2 Practical Realization of Conditional Design

The central idea of conditional design is to introduce constraints into the design process that allow one to quickly discard potential candidates at one node based on conditions imposed by other nodes of the network. Most efficient will be an on-the-fly constraint optimization, in which a suitable set of constraints can be formulated for every node individually and then be applied to the scaffold being optimized. In other words, we define a set of constraints which allows us to stop the local optimization process at one node due to a possible conflict with other nodes, even though the scaffold under consideration is promising for the local optimization step. Clearly, as constraints can represent the requirements of other nodes only approximately, they must not be strictly enforced. For a shell- or generation-wise scaffold construction it will be beneficial to base all constraints on local descriptors of the electronic structure, of which partial charges are a most prominent example.
4.3 Generic Network for Dinitrogen Fixation

We demonstrate conditional design at the example of dinitrogen fixation. The fixation of molecular nitrogen under ambient conditions has a long history in chemistry since the first discovery of a transition-metal complex that can bind the inert dinitrogen molecule in the 1960’s [220]. Nature accomplishes this task within the enzyme nitrogenase [221, 222], which is a clear demonstration of its chemical feasibility, although the detailed mechanism is still far from being understood.

This has stimulated an early, but often unsuccessful intense search for a suitable synthetic dinitrogen-fixation catalysts. One important milestone has been the formulation of a dinitrogen fixation cycle at a single metal center by Chatt in 1976 [78, 79]. However, it was not before 2003 that Yandulov and Schrock were able to present the first catalyst that can accomplish the task at a homogeneous catalyst under ambient conditions [80, 81]. Unfortunately, the turnover number is very low so that the search for improved catalysts continues. As of today, the number of such
catalysts is very small [223–225] and also these other catalysts exhibit very low turnover numbers [226].

While synthetic approaches in this area are resource intensive, computational design could be a solution. The generic catalytic Chatt–Schrock cycle for dinitrogen fixation at a single metal center is depicted in Fig. 4.4. The reason for the low turnover number of the Schrock catalyst is the reduction of the trisamidoamine chelate ligand under reaction conditions, which leads to a decomposition of the catalyst [227]. Hence, side reactions need to be considered at every step of the Chatt–Schrock cycle as the structure of the complex at each node may be subject to decomposition, yielding another layer of complexity to the abstract cycle as depicted in Fig. 4.4.

However, it is possible to condense the feasibility of dinitrogen fixation into four requirements. A successful catalyst must be able to i) coordinate the inert dinitrogen ligand, ii) activate it so that the transfer of the first (two) proton(s) accompanied by one (two) electron(s) is feasible, iii) dissociate the finally produced ammonia ligand (possibly in an exchange reaction against the next incoming dinitrogen molecule) and iv) endure rather drastic reduction and protonation conditions (the strong reductants cobaltocene and chromocene deliver electrons and the strong acid lutidinium the protons).

For our analysis here, we study scaffold design for properties i) and iii) as these two issues are at the heart of dinitrogen fixation. We aim to design a chelate scaffold that allows a metal atom to coordinate dinitrogen sufficiently strongly, while an ammonia ligand at that metal fragment can be exchanged by dinitrogen. This is difficult to accomplish because a $\sigma$-donor ligand like ammonia cannot easily be substituted by the weak $\pi$-acceptor ligand dinitrogen.

A peculiar feature of the Schrock catalyst is that reducing the ammonia complex weakens the coordination bond to ammonia, while at the same time the bond strength of the dinitrogen bond is increased [228]. Moreover, the exchange is likely to occur through a six-coordinate intermediate in an addition–elimination fashion [229]. Hence, the selection of design constraints may either be based on the dinitrogen and ammonia complexes or on the stable six-coordinate intermediate that binds dinitrogen and ammonia at the same time (or on both).

For our design study here, we can thus formulate the conditional design problem as follows: We seek a chelate-ligand scaffold for a molybdenum
center that allows to bind dinitrogen sufficiently strongly, whereas the co-
ordination of ammonia is not too strong so that dinitrogen can eventually
replace ammonia. Scoring functions for the design process are (intrinsic)
coordination energies, for which a scaling relationship for transition
metal complexes was recently found [230], and the reduction of the forces
on all atoms in the GdMC process.

GdMC is an invaluable tool for this design problem as a standard full
structure optimization of imagined complexes will often produce station-
ary points on the potential energy surface, although such compounds
cannot be made. An example would be the full optimization of a five-
coordinated Mo complex with simple ligands such as \([\text{Mo(NH}_3\text{)(NH}_2\text{)}_3\text{N}_2]\)
that is stable upon structure optimization and binds dinitrogen with a
(BP86/def2-TZVP) coordination energy of \(-199.3 \text{ kJ/mol}\). However, such
a complex, if made, can only be stabilized by isolation as it would other-
wise react with an environment. In general, viable ligands will require
larger (chelate) scaffolds.

Within GdMC, one can easily set up a fragment structure [76, 77], in
which dinitrogen is coordinated to Mo. The \(N=N\) internuclear distance
can be fixed to assign a certain degree of activation and the Mo–N dis-
tance can be set based on standard bond-length considerations. Only a
rather complex (chelate) ligand is likely to reduce the gradients on such a
fragment to a sufficient extent.

### 4.4 Conditional Design of Schrock-type Scaffolds

In order to elaborate on the practical realization of conditional design, we
carry out an extensive screening of structures so that all assumptions made
for the design process can be analyzed in detail. For this, we choose a
smaller model of the full Schrock complex in order to establish a screening
without truncation of structure trees for reference. Without this extensive
set of data, it would be impossible to judge whether the defined descriptors
are effective or not.
4.4.1 Preparations for an Extensive Screening of the Schrock-Scaffold Space

We choose the $\ce{N_2}$ node as starting point for our extensive screening, comprised of the $\ce{N_2}$ fragment, the Mo metal center and four binding sites. Before starting the design process, we set specific thresholds for the intrinsic coordination energy $D_{e}^{\text{int}}$ as well as the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ for each generation. Without such thresholds for the scoring functions, the number of structures generated would be enormous and suffer from combinatorial explosion already at a very early stage of the screening process, even by limiting the number of child structures to 1'000. For a detailed explanation of what computational settings were applied here, we refer the reader to Section A.2. In general, these thresholds may be derived based on generic design considerations, particularly if no ligand environment satisfying the desired property is known beforehand. An alternative approach is the inspection of the respective values for all generated structures in a specific generation and defining a cut-off value a posteriori, i.e., such that only a certain percentage of generated structures represents a valid parent structure for the subsequent design step.
In our case, we can take advantage of the Schrock-type catalyst, which is known to be a candidate structure for the network under study and from which we can derive such thresholds straightforwardly. In essence, we want to make sure that we include the structure branch containing the Schrock-type catalyst in our scaffold design process as well as structure candidates with approximately equal or even better values for the scoring functions. Here, we note that not necessarily all structures within such thresholds need to be selected to form parent structures for the next design step. Setting such thresholds merely guarantees that all structures fulfilling these criteria indeed form meaningful parent structures in terms of these two descriptors.

The calculated values for the structure branch containing the Schrock-type catalyst are given in Fig. 4.5. Besides the calculated values for the intrinsic binding energy $D_{\text{int}}^i$ and the fragment gradient $|\nabla_{\text{frag}} E_{\text{cl}}|$, we also present the values for the Löwdin partial charge $q_L$ on the molybdenum atom to which we will come back at a later stage in the design process. Based on the values for the Schrock-type structure branch, we define the following upper bounds for the intrinsic coordination energy $D_{\text{int}}^i$ and the fragment gradient $|\nabla_{\text{frag}} E_{\text{cl}}|$: -180 kJ/mol and 2.50-10^{-2} hartree/bohr for the first generation, -170 kJ/mol and 5.00-10^{-2} hartree/bohr for structures without any chelate topology in the second generation, -180 kJ/mol and 3.00-10^{-2} hartree/bohr for structures with chelate topologies in the second generation, and finally -150 kJ/mol and 1.50-10^{-2} hartree/bohr for the third generation where we start to functionalize candidate structures by benzene derivatives.

We note that all structures in the second generation having at least one chelate arm are separated from all structures in the same generation without any chelate arms. That is, when designing structural candidates for the second generation, we first add atoms and subject these structures to a constraint optimization. In case a structure then lies within the thresholds defined above, we continue the design process with such a structure and form all possible chelate topologies, each resulting in a new structure still belonging to the second generation. We therefore check first if a distinct set of atoms offers a good choice for the ligand environment. If this is not the case, we discard this set of atoms right away, assuming that a possible chelate effect will not be sufficient to make such a structure a good candidate structure. Again, this approach allows us to thin out structure branches as early as possible.
Naturally, we have to make sure that we do not exclude possibly suitable candidate structures by explicitly defining such thresholds for each individual generation of structures. Thus, we randomly probe selected candidate structures in generation n which violate the predefined thresholds and generate new child structures for generation n+1. This procedure allows us to readily identify whether the chosen thresholds are meaningful. Besides probing candidate structures which are slightly above a given threshold, we also probe structures well above a predefined threshold, i.e., arguably irrelevant structures in generation n form parent structures for child structures in generation n+1.

4.4.2 Extensive Screening of the Schrock-Scaffold Space

We now turn to candidate structures for the N₂ node. The values of the intrinsic binding energy \( D^\text{int}_e \) and the fragment gradient \( |\nabla_{\text{frag}} E_{\text{el}}| \) for all generated structures in the first generation are shown in Fig. 4.6. Our automated scaffold designer MSD generates a wide variety of structures with fragment gradients up to 0.10 hartree/bohr and intrinsic binding energies \( D^\text{int}_e \) between -80 kJ/mol and -240 kJ/mol. We visualize the thresholds for the previously defined scoring functions as dotted lines in the diagram. Structure candidates which were selected as parent structures for the continuation of the design process are marked by blue asterisks. Here and in the following diagrams, the structure belonging to the Schrock-type structure branch is marked by a black pentagon. For reasons of feasibility, MSD selected only ten structures within the thresholds for \( D^\text{int}_e \) and \( |\nabla_{\text{frag}} E_{\text{el}}| \) as parent structures for the next iteration of the design process.

In order to justify our predefined thresholds, we randomly pick four structures which do not satisfy the constraints on the scoring functions, labelled as \([2]^4_{N_2}, [3]^4_{N_2}, [4]^4_{N_2}, \) and \([5]^4_{N_2}, \) respectively. The corresponding molecular environments are shown at the bottom of Fig. 4.6. Based on these structures, all possible child structures in the second generation are formed and depicted in the four smaller diagrams, together with the thresholds for the second generation as dotted lines. Particularly for parent structures \([3]^4_{N_2}\) and \([4]^4_{N_2}\), we see that the vast majority of all formed child structures in the second generation does not fulfill the constraints set for this generation. Also, for parent structures \([1]^4_{N_2}\) and \([2]^4_{N_2}\), this is the case for the majority of child structures which shows that the predefined thresholds for the scoring functions are justified, meaning that we do not
Figure 4.6.: **Big Diagram:** Visualization of $D_e^{\text{int}}$ and $|\nabla_{\text{MoN}_2} E_{\text{el}}|$ for all generated structures in the first generation. Structure candidates employed as parent structures for the subsequent design process are marked by a blue asterisk. The structure belonging to the Schrock-type structure branch is marked by a black pentagon. **Small Diagrams:** Visualization of $D_e^{\text{int}}$ and $|\nabla_{\text{MoN}_2} E_{\text{el}}|$ for all generated structures originating from the four highlighted parent structures. The corresponding BP86/def2-TZVP optimized ligand environments are shown in ball-and-sticks representation at the bottom as well as their values for $|\nabla_{\text{MoN}_2} E_{\text{el}}|$ in hartree/bohr and $D_e^{\text{int}}$ in kJ/mol.
Conditional Design of Schrock-type Scaffolds

discard too many possibly relevant structure branches at this early stage of the design process.

For the second generation of structures without chelate topologies, we present the values for the scoring functions for all generated structures in Fig. 4.7. As above, we visualize the thresholds for the previously defined scoring functions as dotted lines and mark the structure candidates selected as parent structures for the continuation of the design process with blue asterisks. From 2'929 possible parent structures which fulfill the constraints on our scoring functions, the MSD selects 664 structures for which all possible chelate substructures are formed. Since all of those structures lie relatively far away from the threshold of 5.00-10^-2 hartree/bohr for |∇_{MoN2}E_{el}|, yet very close to the cutoff for \(D_{e}^{int}\), we only probe the validity of -170 kJ/mol as cut-off value. To this end, we build all possible child structures containing at least one chelate arm out of a set of structures which lies relatively close to the threshold of -170 kJ/mol (indicated by the label '5') and farther away (indicated by the label '6'). The resulting child structures are presented in the two diagrams at the bottom of Fig. 4.7. Again, the vast majority of child structures does not fulfill the constraints (3.00-10^-2 hartree/bohr and -180 kJ/mol) defined for structures containing chelate topologies, that is, these values provide meaningful thresholds.

Based on the selected parent structures in the second generation, the MSD designs all possible child structures containing at least one chelate arm which are presented in Fig. 4.8 on the left hand side. The majority of structures cluster around a value of -180 kJ/mol for \(D_{e}^{int}\) and a value of up to 3.00-10^-2 hartree/bohr for |∇_{MoN2}E_{el}|. For the subsequent design step, the MSD now replaces the outermost hydrogen atoms by functional groups instead of adding more atoms in a consecutive fashion. Influenced by the simplified Schrock-type catalyst where benzene derivatives appear from the third generation onwards, we choose three benzene-type functional groups (benzene, phenol, cumene) for this process which are shown in Fig. 4.9. The values for |∇_{MoN2}E_{el}| and \(D_{e}^{int}\) for the resulting structures are shown on the right hand side in Fig. 4.8. We note that the fragment gradient |∇_{MoN2}E_{el}| for all generated structures in the third generation lies below 3.00-10^-2 hartree/bohr and the value for \(D_{e}^{int}\) for the majority of structures lies between [-190 kJ/mol, -150 kJ/mol].
Chapter 4  |  CONDITIONAL DESIGN OF MOLECULAR SCAFFOLDS

Figure 4.7.: **Big Diagram:** Visualization of $D_{\text{int}}^e$ and $|\nabla_{\text{MoN}_2} E_{\text{el}}|$ for all generated structures in the second generation without chelate topologies. Structure candidates employed as parent structures for the subsequent design process are marked by a blue asterisk. The structure belonging to the Schrock-type structure branch is marked by a black pentagon. **Small Diagrams:** Visualization of $D_{\text{int}}^e$ and $|\nabla_{\text{MoN}_2} E_{\text{el}}|$ for all generated structures originating from the structures within the boxes marked by ‘5’ and ‘6’, respectively. For each subset of structures within those boxes, two representative BP86/def2-TZVP optimized ligand environments were randomly selected and are shown in ball-and-sticks representation at the bottom as well as their values for $|\nabla_{\text{MoN}_2} E_{\text{el}}|$ in hartree/bohr and $D_{\text{int}}^e$ in kJ/mol.

$$\begin{align*}
[6\quad 1]_{\text{N}_2}^4 & \quad (1.41 \cdot 10^{-2}; -164.9) \\
[7\quad 1]_{\text{N}_2}^4 & \quad (1.52 \cdot 10^{-2}; -164.9) \\
[8\quad 1]_{\text{N}_2}^4 & \quad (1.73 \cdot 10^{-2}; -141.7) \\
[9\quad 1]_{\text{N}_2}^4 & \quad (1.58 \cdot 10^{-2}; -140.26)
\end{align*}$$
Figure 4.8.: **Top:** Visualization of $D_{\text{int}}^\text{ex}$ and $|\nabla_{\text{frag}} E_{\text{cl}}|$ for all generated structures in the second generation with chelate topologies (left) and for all generated structures in the third generation (right). Structure candidates employed as parent structures for the subsequent design process are marked by a blue asterisk. The structure belonging to the Schrock-type structure branch is marked by a black pentagon. **Bottom:** For both generations, two representative BP86/def2-TZVP optimized ligand environments were randomly selected and are shown in ball-and-sticks representation at the bottom.

Figure 4.9.: BP86/def2-TZVP optimized functional groups employed for the third generation instead of adding single atoms. From left to right: benzene, phenol, and cumene. The atoms directly connected to the molecular scaffold are marked in green.
Figure 4.10: From left to right: Visualization of $D_{e}^{\text{int}}$ and $|\nabla_{\text{MoNH}} E_e|$ for all generated ligand environments in the first, second, and third generation for the NH$_3$ fragment. The structure belonging to the Schrock-type structure branch is marked by a black pentagon.

4.4.3 Continuation of the Design Process at the NH$_3$ Node

We now continue our design process at the NH$_3$ node, that is, we design a suitable ligand environment which is able to dissociate the ammonia ligand conditioned on the fact that the same ligand environment is able to bind the N$_2$ fragment. As explained in Section 4.1.5, we would only need to subject those molecular scaffolds to a constrained optimization at this node which already fulfilled the thresholds for our scoring functions for the N$_2$ node. However, we subject all structures designed for the N$_2$ node to a constrained optimization for the NH$_3$ node as well in order to obtain a consistent picture of all possible ligand environments. The resulting values for $D_{e}^{\text{int}}$ and $|\nabla_{\text{MoNH}} E_e|$ are shown in Fig. 4.11.
4.4.4 The Löwdin Partial Charge as Valid Conditional Descriptor

As the main goal of our conditional design approach is to thin out structure branches as early as possible in the design process, additional descriptors besides the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ as well as the intrinsic binding energy $D_{\text{int}}^e$ are required to tame the combinatorial explosion. Such descriptors must be local so that structural modifications in the outer periphery do not change the value of the descriptor significantly. If this is not the case and an already cut structure branch in generation $n$ would exhibit a value for a chosen descriptor in generation $n + 1$ which is valid again, the descriptor itself will not be meaningful. Furthermore, as already explained in Section 4.1.6, the definition of such a descriptor and its thresholds should conceptually be straightforward.

One such descriptor is given by the partial charges of the molecule, particularly the partial charge on the transition metal center. In order to avoid basis-set dependence effects, the Löwdin partial charge is preferred over other basis-set dependent partial charges such as the Mulliken partial charge. Naturally, it would be desirable to employ robust conditional descriptors such as the electron density or the electrostatic potential as described by Matta [231, 232], however, this is beyond the scope of this work. We also note that, in general, a conditional descriptor does not have to be of physical nature, i.e., observable in experiment. For us, the most important aspect is the ability of a descriptor to cut additional structure branches as early as possible in the design process.

In order to include a conditional descriptor such as the Löwdin partial charge into our design process, we assume that both intrinsic coordination energy $D_{\text{int}}^e$ and the fragment gradients $|\nabla_{\text{frag}} E_{\text{el}}|$ correlate with the partial charges on the transition metal center. A first analysis towards the inclusion of this descriptor is given in Fig. 4.11 where the dependence of the Löwdin partial charge $qL$ with respect to the intrinsic binding energy $D_{\text{int}}^e$ is depicted.

As a next step, it will now be important to single out those structures which have a low intrinsic binding energy $D_{\text{int}}^e$ in Fig. 4.11, since we aim for a molecular environment which is able to form NH₃. Once those structures are identified, their Löwdin partial charge $qL(N_2)$ at the N₂ node must be calculated in a subsequent step. The result will be a set of structures which fulfill the desired properties at both the N₂ and NH₃
Figure 4.11.: From left to right: Visualization of $D_{e}^{\text{int}}$ and $q^{L}(\text{Mo})$ for all generated ligand environments in the first, second, and third generation for the NH$_3$ fragment. The structure belonging to the Schrock-type structure branch is marked by a black pentagon.

nodes of our reaction network where we also know which boundaries need to be applied for $q^{L}$ in both cases. This in turn will then serve as another descriptor for the generation of additional structure branches during a continuation of the design process.
5

Stereoisomer Discrimination by Chirooptical Vibrational Scattering Spectroscopy

After elaborating on how to construct a molecular scaffold which is able to fulfill a specific property (in the most general case as a component of a network of molecular-rearrangement steps), we now focus on a methodology to identify such complexes by theoretical spectroscopy. To this end, we introduce a certain form of chiroptical vibrational scattering spectroscopy, namely vibrational Raman Optical Activity (VROA) spectroscopy. VROA has become one of the predominant techniques for the determination of the absolute configuration of a molecule [84, 88–90] and for the study of conformational changes of chiral biomolecules [91, 92] besides other chiroptical techniques such as electronic and vibrational circular dichroism (VCD) and optical rotation [86, 87].

While measuring a VROA spectrum is now a routine task, the interpretation of its band structure poses a challenge due to a complex relation between the molecular structure and its spectral pattern. Here, theoretical studies of VROA can be essential for the assignment of observed spectral features to a certain molecular vibration and structure. As an example, we point out several studies by our group which investigated spectroscopic signatures in proteins [233–237]. For some excellent reviews on this topic, we direct the reader to the literature [87, 93–97, 238]. Prominent examples where quantum chemical calculations were combined with experiment to assign an absolute configuration are chirally
deuterated neopentane [239] and methyloxirane in gas and liquid phases [240, 241] as well as the prototypical asymmetric carbon in the molecule CHFClBr [242]. Methodological developments such as density fitting for the calculation of property tensors [243], the tensor transfer method by Bour, Keiderling, and co-workers [244–246], the decomposition into gradient components [247, 248], the interpretation of spectra in terms of localized vibrational modes [249–251], the provision of new graphical tools [252, 253], the access to coupled-cluster data [254], London orbitals and gauge-origin independent spectra [145, 255, 256], surface-enhanced VROA [257, 258], and investigations into how anharmonic effects contribute to a VROA spectrum [259, 260] contributed to the rapid advance of the off-resonance variant of VROA in computational spectroscopy. Even coordination compounds [243, 261, 262] and proteins [263, 264] are now in reach for computational methods.

The VROA intensities are about a thousand times weaker than the already weak Raman intensities. It is therefore desirable to selectively enhance the VROA and Raman signals by adopting a laser excitation wavelength in resonance with an electronic transition. This allows for the enhancement of the Raman signals of, e.g., amide bonds in proteins or aromatic side chains [265–269].

In order to calculate such spectra in resonance, several theoretical developments were accomplished in the last decade. Nafie [270] developed the resonance VROA (RVROA) theory in the limit of a resonance with one excited electronic state. This RVROA theory was later extended by Jensen et al. [145] to include multiple electronic states, via solving for complex linear response tensors by inclusion of a common excited-states damping, which was subsequently implemented in the Amsterdam Density Functional (ADF) program package [271] in a Kohn-Sham (KS) density functional theory (DFT) framework. Our group [272] studied the interference between several excited electronic states for RVROA spectra of naproxen and naproxen-OCD₃ with a sum-over-states (SOS) approach and showed that the consideration of a second excited electronic state may change the RVROA intensities significantly.

Experimental studies showing the RVROA effect for naproxen and ibuprofen were carried out by Vargek et al. [273] and other examples were also reported, see the RVROA study of single-walled carbon nanotube enantiomers by Magg and co-workers [274]. Experimental RVROA spectra involving two electronically excited states for a chiral transition metal
complex were reported by Merten et al. [275]. Very recently, Crassous and co-workers [276] conducted an experimental study where among other techniques VROA spectroscopy was employed to examine modifications upon redox changes to helicene-based mono- and bis-Fe\textsuperscript{II} complexes. The spectra were obtained with a laser wavelength of 532 nm and evidently under resonance, as they were mono-signate and followed the intensity patterns of the corresponding resonance-Raman spectra.

It is obvious that computational methods for RVROA intensities should consider many electronically excited states. Here, we describe such an implementation which is based on the short-time approximation to the time-dependent theory of Raman scattering as formulated by Heller et al. [277–279]. In the off-resonance case, it is identical to the standard theory introduced by Placzek, while it becomes a Placzek-like polarizability theory in the resonance case. If only short-time dynamics is involved, that is, the excited states dephase rapidly compared to the time scale of molecular vibrations, this theory is accurate. We apply a (semi-) numerical differentiation scheme for the evaluation of the spectroscopic intensities as numerical first derivates of the respective property tensors, which makes our implementation easily adaptable to new electronic structure methods that can deliver such property tensors, whereas their analytic derivatives might be more difficult to derive and implement. The numerical errors introduced are usually negligible in the framework of the double harmonic approximation [148, 280]. Hence, this approach has two advantages: First, one needs only analytic geometry gradients and molecular property tensors, and not analytic Hessian and property gradients, respectively, for the electronic structure method of choice, and secondly, it parallelizes trivially and makes large molecules easily accessible.

This methodology we use to study the resonance and off-resonance VROA spectra of H\textsubscript{2}O\textsubscript{2}, (S)-methyloxirane, (S)-naproxen, and (S)-ibuprofen in this chapter. Our goal is, first, the validation of our implementation, and second, the assignment of band patterns found in experiment to molecular vibrations and the interpretation of experimental observations in terms of our calculated results.
5.1 VROA and RVROA of H₂O₂

Even though H₂O₂ is chiral in its equilibrium structure in the Born-Oppenheimer approximation, torsional tunneling between the two enantiomeric forms occurs which is why this chirality cannot be resolved experimentally. Yet, due to its small size, H₂O₂ (and also methyloxirane in Section 5.2) is a popular benchmark system for quantum chemical calculations of chiroptical properties of molecules [87, 145, 281–286].

An important aspect of the RVROA calculations is the ability to reproduce the optical rotatory dispersion around and at resonance correctly. Fig. 5.1 shows the molar ellipticity, i.e., the electronic circular dichroism (ECD) spectrum, of H₂O₂ calculated in two different ways: In one case, the imaginary part of the optical rotation parameter $\beta = -G'(\omega)/\omega$ from the complex damped response calculations was converted to the ECD intensity on a grid of frequency points. The curve exhibits the Lorentzian damping intrinsic in the approach. For comparison, the excitation energies and rotatory strengths were calculated with the KS (TDDFT) excitation module of NWChem and Lorentzian-broadened subsequently. The same Gaussian-type basis sets (def2-TZVP) and functionals (BLYP) were used in both cases. For the damping parameter $\Gamma$, a value of $\Gamma = 0.0037$ a.u. was applied. After test calculations showed that using a GIAO basis is not
The spectral lines calculated with BLYP/def-TZVP are indicated by the blue color, for BLYP/def-QZVP by the orange color. For the calculated VROA spectrum, we adopted an incident wavelength of 514.5 nm, whereas for the RVROA spectrum, we chose a wavelength of 241.3 nm and 243.4 nm for the calculations with the def2-TZVP and def2-QZVP basis set, respectively. Calculated spectra were broadened with a Lorentzian having a FWHM of 20 cm$^{-1}$.

vital as long as the molecules are not very large, as is the case here, and centered at the coordinate origin, the dipole-length gauge was chosen as it entails computational savings. The two ECD curves are seen to match perfectly.

The calculated VROA and RVROA spectra of H$_2$O$_2$ employing two different basis sets are presented in Fig. 5.2 for wave numbers below 1800 cm$^{-1}$. For the def2-TZVP basis set, the RVROA spectrum is calculated at 241.3 nm (with an oscillator strength of $F = 1.55 \times 10^{-3}$ and a rotatory strength of $R = -7.97 \times 10^{-40}$ esu$^2$ cm$^2$), whereas for the def2-QZVP basis set, a wavelength of 243.4 nm ($F = 2.92 \times 10^{-3}$, $R = -10.03 \times 10^{-40}$ esu$^2$ cm$^2$) is adopted, both corresponding to the lowest electronic transition calculated with BLYP and the respective basis set. As seen in Fig. 5.2, the VROA spectrum is characterized by two positive and two negative bands.
For the calculated VROA spectrum on the left hand side, the two positive bands at 349 cm\(^{-1}\) and 1245 cm\(^{-1}\) correspond to the torsional vibration and the asymmetric O-H bending mode, respectively. The two negative bands at 857 cm\(^{-1}\) and 1369 cm\(^{-1}\) represent the O-O stretching mode and the symmetric O-H bending mode, respectively. The RVROA spectrum on the right hand side in Fig. 5.2 shows only three positive bands. The monosignate character of the spectrum is in good agreement with the two-state approximation [270] which states that the RVROA intensities are monosignate with an opposite sign of the CD signal at the same wavelength. A calculated rotatory strength of \(-7.97 \times 10^{-40}\) esu\(^2\) cm\(^2\) and \(-10.03 \times 10^{-40}\) esu\(^2\) cm\(^2\) employing the length gauge with the def2-TZVP and the def2-QZVP basis set, respectively, yields the observed positive signature. Furthermore, the asymmetric O-H bending mode at 1245 cm\(^{-1}\) is not enhanced which is also in agreement with resonance Raman scattering (RRS) where normally only the symmetric vibrations are strong enhanced. Comparing the intensities of the VROA spectrum with the RVROA spectrum, we see a resonance enhancement on the order of \(10^4\), which can be typically found in RRS as well [287]. We also show the VROA and RVROA spectra of H\(_2\)O\(_2\) on top in Fig. 5.2 as calculated by Jensen and co-workers [145]. Both spectra compare well with our calculations with respect to the position of the normal modes as well as the spectral intensities.

In Table 5.1, we present the backscattering VROA intensities calculated at 514.5 nm, employing BLYP/def2-TZVP. We compare the numerical values of \(\beta(G')^2\) and \(\beta(A)^2\) to the ones obtained by a MoVIpAC-TURBOMOLE calculation. To allow for this comparison, we first made sure that both codes returned exactly the same SCF energies for one specific geometry of H\(_2\)O\(_2\). We then employed the def2-TZVP and def2-QZVP basis sets with spherical basis functions as well as the BLYP functional and the dipole-length gauge in both cases. For the numerical integration of the exchange-correlation potential, we chose comparable grids (m4 for TURBOMOLE, xfine for NWChem). Since TURBOMOLE does not offer the functionality to include contributions from the excited states to the property tensors, we chose the off-resonance functionality of NWChem for better comparison. The signs and, in general, the magnitudes of the property tensors are in agreement with the MoVIpAC-TURBOMOLE calculation.
Table 5.1: Backscattering VROA intensities at 514.5 nm in 10^{-6} \text{ Å}^4 \text{ amu}^{-1} employing the BLYP density functional and def2-TZVP basis set. The VROA intensities are reported without the numerical factors as $\beta(G')^2_\rho$ and $\beta(A)^2_\rho$.

5.2 VROA and RVROA of (S)-methyloxirane

For (S)-methyloxirane we compare our VROA results with those published by our group before [286]. The VROA differential cross sections in the gas phase are presented in Fig. 5.3 for wave numbers below 1800 cm$^{-1}$. Our calculated spectrum is in good agreement with the results which we obtained in our former study [286], both for the sign pattern of the spectrum as well as the magnitude of the intensities. Particularly the positive peak at 721 cm$^{-1}$ followed by the two negative peaks at 799 cm$^{-1}$ and 877 cm$^{-1}$ is reproduced very well. However, there are also a few differences between the experimental VROA spectrum and our calculated VROA spectrum, particularly for the low-frequency mode at 399 cm$^{-1}$, where our calculations yield a different sign of the intensity. Since this mode is weak, we expect a high sensitivity to the employed basis set and functional. It should be noted that the VROA intensities in backscattering geometry within the MoVíPac software suite are defined as [288]:

$$I^R (180^\circ) - I^L (180^\circ) = \kappa_{\text{MoVíPac}} \frac{4}{c} \left( 24\beta(G')_\rho + 8\beta(A)_\rho \right), \quad (5.1)$$

whereas in ADF, the following definition holds [145]:

\[ \]
Chapter 5 | STEREOISOMER DISCRIMINATION BY RVROA

Figure 5.3.: Calculated VROA differential cross sections for (S)-methyloxirane in units of Å²/amu⁻¹ at an incident wavelength of 488.9 nm. **Top left:** Experimental backscattering VROA spectrum measured by Hug and reported in Ref. [286]. **Top right:** Reproduction of the calculated VROA spectrum employing BLYP/TZP and GIAOs obtained with ADF in Ref. [145]. **Bottom:** Our calculated VROA spectrum employing BLYP/def2-TZVP. All calculated spectra were broadened with a Lorentzian having a FWHM of 20 cm⁻¹.

\[ I^R (180^\circ) - I^L (180^\circ) = \kappa_{ADF} \left[ \frac{48(\beta(G')_p + \beta(A)_p/3)}{90c} \right]. \tag{5.2} \]

Here, the parameters \( \kappa_{ADF} \) and \( \kappa_{MoViPac} \) are defined differently such that Eq. 5.1 differs by a factor of 2 compared to Eq. (5.2) (see also a related comment in Ref. [145] in Section B). Keeping this in mind, our calculated spectrum compares well with the experimental spectrum, both with respect to the magnitude of the intensities and the overall sign pattern.

For the RVROA spectra, we calculate the two lowest excited states with BLYP/def2-TZVP at 184.0 nm (\( F = 1.194 \times 10^{-2}, R = 28.84 \times 10^{-40} \) esu² cm²) and at 169.4 nm (\( F = 3.9 \times 10^{-2}, R = -1.22 \times 10^{-40} \) esu² cm²). These compare
VROA and RVROA of (S)-methyloxirane

Figure 5.4.: Top: Reproduction of the calculated RVROA spectra for methyloxirane employing BLYP/TZP, $\Gamma=0.0037$ a.u., and modified velocity gauge (MVG) obtained with ADF in Ref. [145]. Middle and bottom: Calculated RVROA and Raman differential cross sections for (S)-methyloxirane in units of $\text{Å}^2/\text{amu}^{-1}$ for the $S_1$ state (left) and the $S_2$ state (right) at an incident wavelength of 184 nm and 169 nm, respectively. All spectra were calculated employing BLYP/def2-TZVP, $\Gamma=0.0037$ and a Lorentzian broadening with FWHM of 20 cm$^{-1}$ was applied.
well with the transitions found experimentally at 174 nm and 160 nm, respectively [289]. The calculated RVROA spectra for the S₁ and S₂ excited states are presented in Fig. 5.4. Based on the two-state approximation, we expect a negatively monosignate spectrum and a positively monosignate spectrum for the S₁ and S₂ excited state, respectively, due to the corresponding signs of the rotatory strength. This is in fact the case and we observe an enhancement on the order of 10⁴. We also show the RVROA spectra of methyloxirane as calculated by Jensen and co-workers [145] for the first and second excited state on top in Fig. 5.4. Particularly for the first excited state, our calculated RVROA spectrum compares well with their RVROA spectrum.

For the second excited state, we do not find a completely monosignate spectrum for the calculated wavelength corresponding to the S₂ state, as is the case for the RVROA spectrum of Jensen and co-workers [145]. Our spectrum shows both positive and negative peaks. By scanning over the wavelength range from 160-190 nm we are able to recover a wavelength of 167 nm where the RVROA spectrum shows the typical monosignature as presented in Fig. 5.5 (this wavenumber corresponds to the third excited state S₃ with \( F = 4.183 \times 10^{-2} \), \( R = -26.25 \times 10^{-40} \) esu² cm²). This technique is also adopted in experiment by tuning the wavelength of the laser in order to selectively enhance different chromophores [290, 291].

Another important prediction made by the two-state approximation is the fact that the RVROA spectrum should be identical to the corresponding RRS spectrum, but with a sign determined by the opposite sign of the rotatory strength at the excitation wavelength. For the S₁, S₂, and S₃ excited states, we present the corresponding Raman differential cross
sections in Figs. 5.4 and 5.5. Again, the calculated RVROA spectra are in qualitative agreement with the corresponding Raman spectra.

### 5.3 VROA and RVROA Spectra of (S)-Naproxen and (S)-Ibuprofen

One of the first experimental RVROA studies, by Vargek et al. [292], recorded the VROA and Raman spectra of (S)-naproxen (shown in Fig. 5.7) and compared with the results for (S)-ibuprofen. The optimized molecular structures for both molecules are depicted in Fig. 5.6. The study by Vargek et al. revealed a resonance enhancement of the spectra for (S)-naproxen at an incident laser wavelength of 514.5 nm, indicated by an absorption tail. This arose from a strong band in the electronic absorption spectrum around 325 nm. By contrast, (S)-ibuprofen is characterized by a typical off-resonance VROA spectrum at the same laser wavelength.

We first investigate the VROA spectra of (S)-naproxen and (S)-ibuprofen calculated at an incident laser wavelength of 514.5 nm. Both calculated VROA spectra are shown in Fig. 5.8. The VROA spectrum of naproxen is mainly characterized by positive peaks across the wavelength range from 300 cm\(^{-1}\) to 1800 cm\(^{-1}\). We identify the negative peak at 1292 cm\(^{-1}\) corre-
Figure 5.7.: Top: Raman and VROA spectra of (S)-naproxen. The spectra were recorded with 500 mW of 514.5 nm laser excitation and the sample was dissolved in 80 mg/ml NaOH, pH 10. Figure reproduced from data published in Ref. [292]).

Bottom: Raman and VROA spectra of (S)-ibuprofen. The spectra were recorded with 500 mW of 514.5 nm laser excitation, and the sample was dissolved in CCl₄ (0.21 M). Figure reproduced from data published in Ref. [273].
sponding to the O–H bending vibration of the carboxyl group and to the C\textsubscript{α}–H bending vibration of the adjacent carbon atom. The strongest positive peak at 1602 cm\textsuperscript{-1} is caused by C–H stretching and bending vibrations of the naphthalene ring framework (i.e., a graphene-like C–C stretching mode). Compared to the experimental spectrum of (S)-naproxen in Fig. 5.7, we identify this strong VROA-active band with the experimental peak at around 1400 cm\textsuperscript{-1}, that is, shifted to lower wavenumbers. This shift could be caused by the experimental conditions, as the experiment was conducted in NaOH solution. For now, we neglect solvation effects as we intend to provide calculated spectra for well-defined molecular structures, which is the reason why we chose the uncharged gas-phase conformers depicted in Fig. 5.6. In Section 5.4, we study in detail possible solvation effects on the calculated spectra by means of microsolvation.

Another strong positive peak appears at 1722 cm\textsuperscript{-1} corresponding to the C=O stretching vibration of the carboxyl group. Both of these peaks lead to the overall strong positive signature in the upper range of the VROA spectrum. The C–C stretching vibrations are also responsible for the weakly positive band at 1539 cm\textsuperscript{-1}, whereas the positive peak at 1582 cm\textsuperscript{-1} can be attributed to C–H bending vibrations. We identify two weak negative peaks at 1464 cm\textsuperscript{-1} and 1470 cm\textsuperscript{-1}. However, due to the strong positive signal of the normal mode at 1465 cm\textsuperscript{-1}, this -/+/- couplet resolves in a positive band pattern as well (all made up of C–H bending vibrations of both terminal CH\textsubscript{3} groups). Nonetheless, there are some smaller negative bands, such as the one from 946 cm\textsuperscript{-1} to 1018 cm\textsuperscript{-1} which are mainly comprised of localized C–H stretching and bending vibrations of C–H bonds.

While the VROA spectrum of (S)-naproxen clearly shows a resonance enhancement effect, as almost all bands carry a positive signature, the calculated VROA spectrum of (S)-ibuprofen at the bottom of Fig. 5.8 is characterized by a typical off-resonance signal pattern with both positive and negative bands. Furthermore, the bands in the VROA spectrum are an order of magnitude weaker compared to the calculated VROA spectrum of (S)-naproxen.

In contrast to the calculated VROA spectrum of (S)-naproxen, the two modes at 1587 cm\textsuperscript{-1} and 1721 cm\textsuperscript{-1} which correspond to the C–H bending vibrations of the phenyl ring system and C=O stretching vibration of the carboxyl group, respectively, are characterized by a negative signature. Yet as before, the graphene-like C–C stretching mode of the inner phenyl
Figure 5.8.: Calculated VROA differential cross sections for naproxen (top) and ibuprofen (bottom) in units of Å$^4$/amu$^{-1}$ at an incident wavelength of 514.5 nm. All spectra were calculated employing BLYP/def2-TZVP and a Lorentzian broadening with FWHM of 20 cm$^{-1}$ was applied.
ring at 1587 cm\(^{-1}\) is strongly VROA-active as well. Several couplets dictate the overall behavior of the spectral band, e.g. a strong -/+ couplet at 1460 cm\(^{-1}\). This spectral region corresponds to C–H stretching and twisting vibrations of the CH\(_3\) groups. The strongest positive peak in the range from 0 cm\(^{-1}\) to 1800 cm\(^{-1}\) arises from C–H bending vibrations across the whole molecular structure at 1323 cm\(^{-1}\). Due to two adjacent negative peaks, particularly at 1312 cm\(^{-1}\), this resolves into a -/-/+/- couplet instead of leading to a strong positive band. Except for the negative peak at 1721 cm\(^{-1}\), i.e., the one corresponding to the C=O stretching vibration of the carboxyl group, our calculated spectrum compares well with the experimental spectrum shown in Fig. 5.7.

We now analyse which particular bands are enhanced by means of resonance VROA. We therefore calculate the RVROA spectra under full resonance conditions, viz. at an incident wavelength corresponding to the first excited state \(S_1\). For (S)-naproxen, we calculate this transition with BLYP/def2-TZVP at 355.96 nm (with an oscillator strength of \(F = 4.959 \times 10^{-2}\) and a rotatory strength of \(R = -8.0229102 \times 10^{-40}\) esu\(^2\) cm\(^2\)) and for (S)-ibuprofen with BLYP/def2-TZVP at 265.95 nm (\(F = 8.630 \times 10^{-2}\), \(R = -21.4772644 \times 10^{-40}\) esu\(^2\) cm\(^2\)). The corresponding RVROA spectra for both molecules are shown in Fig. 5.9.

For (S)-naproxen, the intensity of the overall spectral intensity increases on the order of \(10^3\) compared to the near-resonance VROA spectrum discussed above. We identify three bands that are strongly RVROA-active, namely the bands with wavenumbers 1344 cm\(^{-1}\), 1360 cm\(^{-1}\), and 1539 cm\(^{-1}\). The corresponding normal modes are depicted from left to right in the upper part of Fig. 5.10. All three modes are characterized by C–H bending vibrations within the naphthyl ring system, the strongest one at 1539 cm\(^{-1}\). Due to the fact that two of these normal modes (1344 cm\(^{-1}\) and 1360 cm\(^{-1}\)) lie very closely to each other, they form a strong positive band pattern. In contrast to the VROA spectrum, where we obtained a strong positive band for the normal mode at 1722 cm\(^{-1}\) for the C=O stretching vibration of the carboxyl group and at 1602 cm\(^{-1}\) for the C–H stretching and bending vibrations of the naphthalene ring framework, these two modes show almost no intensity in the RVROA spectrum.

The positively monosignate character of the spectrum is in agreement with the two-state approximation [270] which states that the RVROA intensities are monosignate with an opposite sign of the circular dichroism (CD) signal at the same wavelength. When comparing the RVROA
spectral band with the Raman spectrum calculated at the same incident wavelength of 355.96 nm, an almost identical band pattern appears in both cases (see the bottom left in Fig. 5.9). The only band pattern which deviates in the Raman spectrum, belongs to the normal mode at 517 cm\(^{-1}\) which is mainly comprised of scaffold vibrations of the phenyl ring system and results in a relatively larger peak in the Raman spectrum.

For the RVROA spectrum of (S)-ibuprofen, the two-state model predicts a negatively monosignate character based on the positive sign of the CD signal at the incident wavelength. Indeed, our calculations produce a completely monosignate pattern with negative sign, as shown in Fig. 5.9 on the right hand side. In particular, we predict strong RVROA activity for the C=O stretching vibration at 1721 cm\(^{-1}\) in addition to another strong RVROA band at 1587 cm\(^{-1}\) corresponding to C–H bending vibrations within the phenylene ring. These two normal modes are visualized on the bottom in Fig. 5.10. The corresponding Raman spectrum shown in Fig. 5.9 yields an identical band pattern compared to the RVROA spectrum, but – of course – with positive sign.

Figure 5.9: Calculated RVROA and Raman differential cross sections for (S)-naproxen (left) and (S)-ibuprofen (right) in units of Å\(^2\)/amu\(^{-1}\) at an incident wavelength of 356 nm and 265.9 nm, respectively. All spectra were calculated employing BLYP/def2-TZVP, \(\Gamma = 0.0037\), and a Lorentzian broadening with FWHM of 20 cm\(^{-1}\) was applied.
Figure 5.10.: Normal modes with the strongest RVROA activity for the $S_1$ excited state for (S)-naproxen (top and middle) and (S)-ibuprofen (bottom).
Table 5.3: Oscillator and rotatory strengths for naproxen (left) and ibuprofen (right) for the first ten single excitations. Energies are given in eV, oscillator strengths are unitless, and rotatory strengths in $1 \times 10^{-16}$ esu cm.$^{-2}$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Naproxen</td>
<td></td>
<td></td>
<td></td>
<td>Ibuprofen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Single Excitations
5.4 VROA and RVROA for Micro-Solvated Naproxen and Ibuprofen Carboxylate Anions

Solvation effects are very difficult to model reliably for solvents that explicitly interact through directed bonding toward the solute (e.g., hydrogen bonds); see, e.g., Refs. [237, 293–299]. In the case of naproxen and ibuprofen, hydrogen bonding with surrounding water molecules will mostly affect the methoxy and the carboxyl groups, and therefore not all vibrations of these molecules. Note also that conformational changes might affect some bands of the calculated VROA spectra, which we, expect to be small, considering the rigid aromatic core structures separating the substituents that can occur in different conformations. Still, in order to understand changes in the spectra upon deprotonation and solvation, we consider minimally microsolvated models. We deprotonate the carboxyl group of naproxen and ibuprofen and place one water molecule adjacent to each carboxylate group. The resulting structure is then fully optimized and subjected to spectra calculations. Clearly, this crude minimal microsolvated model that lacks electrostatic embedding, which is particularly important for this anionic species, cannot be taken as a suitable model for the actual conditions present in experiment. However, we conduct this study to understand the change in the spectra for a worst case scenario.

Besides calculating the VROA spectra at the incident wavelength of 514.5 nm, which was chosen in experiment, we also calculate the RVROA spectra at the incident wavelength corresponding to the first electronically excited state for each anion. To this end, we first calculate the energies of the ten lowest electronically excited states after fully optimizing each anion with its water molecule. The corresponding excitation energies are given in Table 5.4 for the naproxen carboxylate anion and in Table 5.5 for the ibuprofen carboxylate anion.

While the first electronically excited state of the ibuprofen carboxylate anion lies at 430.8 nm and therefore far away from the incident wavelength of 514.5 nm, we find that for the naproxen carboxylate anion the wavelength of 514.5 nm lies very closely to the wavelength corresponding to the third electronically excited state, i.e., 506.6 nm. We therefore conclude that for the calculated VROA spectrum for the naproxen carboxylate anion at 514.5 nm the contribution of the excited states to the
Table 5.4: Oscillator and rotatory strengths for the naproxen carboxylate anion with one adjacent water molecule for the first ten singlet excitations. Energies are given in eV, oscillator strengths are unitless, and rotatory strengths in $1 \times 10^{-10}$ esu$^2$ cm$^2$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.8425</td>
<td>$3.71 \times 10^{-3}$</td>
<td>-8.460</td>
</tr>
<tr>
<td>2</td>
<td>2.0023</td>
<td>$2.42 \times 10^{-2}$</td>
<td>-19.181</td>
</tr>
<tr>
<td>3</td>
<td>2.4474</td>
<td>$2.69 \times 10^{-3}$</td>
<td>0.548</td>
</tr>
<tr>
<td>4</td>
<td>2.6242</td>
<td>$2.0 \times 10^{-5}$</td>
<td>0.013</td>
</tr>
<tr>
<td>5</td>
<td>2.8697</td>
<td>$1.79 \times 10^{-2}$</td>
<td>-9.744</td>
</tr>
<tr>
<td>6</td>
<td>2.9969</td>
<td>$9.19 \times 10^{-2}$</td>
<td>4.082</td>
</tr>
<tr>
<td>7</td>
<td>3.3171</td>
<td>$2.1 \times 10^{-4}$</td>
<td>-0.033</td>
</tr>
<tr>
<td>8</td>
<td>3.4089</td>
<td>$1.8 \times 10^{-4}$</td>
<td>-0.003</td>
</tr>
<tr>
<td>9</td>
<td>3.4299</td>
<td>$6.0 \times 10^{-5}$</td>
<td>-0.003</td>
</tr>
<tr>
<td>10</td>
<td>3.4373</td>
<td>$7.2 \times 10^{-4}$</td>
<td>0.133</td>
</tr>
</tbody>
</table>

property tensors have to be accounted for, i.e., we calculate a RVROA spectrum at 514.5 nm by adopting a damping of $\Gamma = 0.0037$ a.u.

5.4.1 RVROA for the Naproxen Carboxylate Anion

We compare the RVROA spectrum at 514.5 nm for the naproxen carboxylate anion with the corresponding VROA spectrum of naproxen at 514.5 nm in Fig. 5.11. Our focus is here on the VROA bands which were discussed in Section 5.3 and we compare with the corresponding RVROA bands of the naproxen carboxylate anion.

The strongest positive peak at 1602 cm$^{-1}$ in the VROA spectrum of naproxen caused by C–H stretching and bending vibrations of the naphthalene ring framework is now found at 1595 cm$^{-1}$ in the RVROA spectrum of the naproxen carboxylate anion, i.e., slightly shifted to lower wavenumbers. However, also here it is the strongest positive band in the RVROA spectrum. The other strong positive peak at 1722 cm$^{-1}$ in the VROA spectrum of naproxen corresponding to C=O stretching vibrations of the carboxyl group can be found at 1584 cm$^{-1}$ for the naproxen carboxylate anion, also here as a strong positive peak. The weakly positive band at 1539 cm$^{-1}$ in the VROA spectrum of naproxen made up of C–C stretching vibrations is now located at 1532 cm$^{-1}$, also as a positive band.
The positive peak at 1582 cm$^{-1}$ of the VROA spectrum of naproxen for C–H bending vibrations is also shifted to lower wavenumbers, that is, to 1577 cm$^{-1}$. We note that the range between 1100 cm$^{-1}$ and 1300 cm$^{-1}$ in the RVROA spectrum of the naproxen carboxylate anion shows only negative RVROA intensities, which is in contrast to the VROA spectrum of naproxen. The most negative peak results for a C–H bending vibrations and symmetric stretching vibrations of the carboxylate group.

For the RVROA spectrum of naproxen corresponding to the first electronically excited state, i.e., at an incident wavelength of 356.0 nm, we identified three strong RVROA-active bands at 1344 cm$^{-1}$, at 1360 cm$^{-1}$, and at 1539 cm$^{-1}$. For the naproxen carboxylate anion, we calculate this transition with BLYP/def2-TZVP at 672.9 nm (with an oscillator strength of $F = 3.71 \times 10^{-3}$ and a rotatory strength of $R = -8.460 \times 10^{-40}$ esu$^2$ cm$^2$, see Table 5.4). The corresponding normal modes in the RVROA spec-
Figure 5.12.: Top: Calculated RVROA differential cross sections for naproxen in units of $\text{Å}^4$/amu$^{-1}$ at an incident wavelength of 356.0 nm corresponding to the transition to the first electronically excited state. Bottom: Calculated RVROA differential cross sections for the naproxen carboxylate anion with one adjacent water molecule in units of $\text{Å}^4$/amu$^{-1}$ at an incident wavelength of 672.9 nm. Both spectra were calculated employing BLYP/def2-TZVP and a Lorentzian broadening with FWHM of 20 cm$^{-1}$ was applied.

The spectrum can be found at 1342 cm$^{-1}$, at 1358 cm$^{-1}$, and at 1532 cm$^{-1}$, respectively, hence all shifted to lower wavenumbers. However, in the RVROA spectrum of naproxen, the normal mode at 1539 cm$^{-1}$ with the strongest RVROA intensity was found for C–H bending vibrations in the naphthyl ring system (see also Fig. 5.13 on the left-hand side). By contrast, the RVROA spectrum of the naproxen carboxylate anion shows only one very strong RVROA band corresponding to the normal mode at 1584 cm$^{-1}$ which is characterized by the C=O stretching vibration of the carboxylate group. Unlike in the RVROA spectrum of naproxen, the normal modes corresponding to C–H bending vibrations are not resonance enhanced. The positively monosignate character of the RVROA spectrum of the naproxen carboxylate anion is in agreement with the two-state approximation.
Figure 5.13: Normal modes with the strongest RVROA activity for the first electronically excited state in the case of naproxen (left) and the naproxen carboxylate anion (right).

5.4.2 VROA/RVROA for the Ibuprofen Carboxylate Anion

In analogy to the discussion above, we compare the VROA spectrum at 514.5 nm for the ibuprofen carboxylate anion with the corresponding VROA spectrum of ibuprofen at 514.5 nm in Fig. 5.14. Again, we focus on the VROA bands which are discussed in detail in Section 5.3 and compare with the corresponding VROA bands of the ibuprofen carboxylate anion. As was the case in the RVROA spectrum of the naproxen carboxylate anion, we observe a shift to lower wavenumbers of the normal modes of interest. For instance, the C–H bending vibrations of the phenyl ring, i.e., the graphene-like C–C stretching mode, at 1587 cm\(^{-1}\) with negative VROA intensity for ibuprofen can now be found at 1581 cm\(^{-1}\) in the VROA spectrum of the ibuprofen carboxylate anion, but with a strong positive VROA intensity. Furthermore, the C=O stretching vibration which was at 1721 cm\(^{-1}\) with negative VROA intensity for the uncharged generic molecule is now at 1579 cm\(^{-1}\) in the VROA spectrum of the ibuprofen carboxylate anion, yet with strongly positive VROA intensity. Together, these two closely-lying modes cause a strongly positive VROA band. Furthermore, the H\(_2\)O molecule with its bending vibrations causes an additional positive peak at 1679 cm\(^{-1}\) in the VROA spectrum of the ibuprofen carboxylate anion.

We identified the strongest positive peak in the VROA spectrum of ibuprofen at 1323 cm\(^{-1}\) for C–H bending vibrations delocalized across the whole molecular structure. The corresponding normal mode in the VROA
Table 5.5: Oscillator and rotatory strengths for the ibuprofen carboxylate anion with one adjacent water molecule for the first ten singlet electronic excitations. Energies are given in eV, oscillator strengths are unitless, and rotatory strengths in $1 \times 10^{-40}$ esu$^2$ cm$^2$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8780</td>
<td>$4.72 \times 10^{-3}$</td>
<td>1.347</td>
</tr>
<tr>
<td>2</td>
<td>2.9390</td>
<td>$2.45 \times 10^{-4}$</td>
<td>-5.883</td>
</tr>
<tr>
<td>3</td>
<td>2.9645</td>
<td>$1.07 \times 10^{-2}$</td>
<td>5.256</td>
</tr>
<tr>
<td>4</td>
<td>3.1222</td>
<td>$7.69 \times 10^{-2}$</td>
<td>-13.988</td>
</tr>
<tr>
<td>5</td>
<td>3.3045</td>
<td>$9.30 \times 10^{-3}$</td>
<td>-1.352</td>
</tr>
<tr>
<td>6</td>
<td>3.3907</td>
<td>$5.21 \times 10^{-3}$</td>
<td>1.120</td>
</tr>
<tr>
<td>7</td>
<td>3.4709</td>
<td>$6.90 \times 10^{-3}$</td>
<td>-2.020</td>
</tr>
<tr>
<td>8</td>
<td>3.5464</td>
<td>$1.16 \times 10^{-2}$</td>
<td>2.498</td>
</tr>
<tr>
<td>9</td>
<td>3.6567</td>
<td>$5.5 \times 10^{-4}$</td>
<td>-0.0490</td>
</tr>
<tr>
<td>10</td>
<td>3.7294</td>
<td>$1.6 \times 10^{-4}$</td>
<td>0.0493</td>
</tr>
</tbody>
</table>

The spectrum of the ibuprofen carboxylate anion is now be found at 1324 cm$^{-1}$, i.e., at an almost identical position, yet with negative VROA intensity. We also note that the C-H bending vibrations in the naphthyl ring system at 1538 cm$^{-1}$ with the most negative VROA intensity for ibuprofen are now positioned at 1550 cm$^{-1}$ with virtually no VROA intensity.

We now turn to the RVROA spectrum of the ibuprofen carboxylate anion which we calculated at an incident wavelength of 430.8 nm. This wavelength corresponds to the transition to the first electronically excited state (with an oscillator strength of $F = 4.72 \times 10^{-3}$ and a rotatory strength of $R = 1.347 \times 10^{-40}$ esu$^2$ cm$^2$ calculated with BLYP/def2-TZVP). Based on the two-state approximation, we would expect a negatively monosignate character, yet we observe several positive bands around 800 cm$^{-1}$ (C–H out-of-plane vibrations) as well as between 1100 cm$^{-1}$ and 1300 cm$^{-1}$ (C–H bending vibrations across the whole molecular structure). Analogous to the RVROA spectrum of ibuprofen, the graphene-like C–C stretching vibration in the RVROA spectrum of the ibuprofen carboxylate anion at 1579 cm$^{-1}$ is resonance enhanced with the strongest negative RVROA intensity.

We present the normal modes with the strongest RVROA intensities in Fig. 5.16 for ibuprofen as well as ibuprofen carboxylate anion.
Figure 5.14.: Calculated VROA differential cross sections for the ibuprofen carboxylate anion with one adjacent water molecule in units of Å$^4$/amu$^{-1}$ at an incident wavelength of 514.5 nm. The spectrum was calculated employing BLYP/def2-TZVP and a Lorentzian broadening with FWHM of 20 cm$^{-1}$ was applied.
Figure 5.15.: Calculated RVROA differential cross sections for the ibuprofen carboxylate anion with one adjacent water molecule in units of Å^4/amu^{-1} at an incident wavelength of 430.8 nm. The spectrum was calculated employing BLYP/def2-TZVP and a Lorentzian broadening with FWHM of 20 cm^{-1} was applied.

Figure 5.16.: Normal modes with the strongest RVROA activity for the first electronically excited state in the case of ibuprofen (left) and the ibuprofen carboxylate anion (right).
First Application of RVROA to Chiral Metal Complexes

After having benchmarked our RVROA implementation on some smaller molecular systems, we now turn to a more challenging case, that is, calculating resonance-enhanced VROA spectra for a chiral metal complex. Recent progress in organometallic chemistry has lead to the discovery of novel chiral systems which exhibit strong chiroptical properties [300–302]. For example, certain aryl-alkynyl-iron systems allow for the redox triggering of their optical properties by means of a facile electron transfer [303, 304] which is a sought for property in many optoelectronic applications [305, 306].

In a recent study, Crassous and co-workers [276] introduced ethynyl-helicene mono- and bis-FeII complexes grafted with one or two electroactive [Fe(k2-dppe)(η5-C5Me3)] fragments [307] exhibiting unprecedented redox-triggered chiroptical switching. They investigated the changes of the vibrational modes upon oxidation by means of vibrational circular dichroism and Raman optical activity and found that the sign of the optical rotation changes remarkably at 1.54 μm while the topology of the helicene is unaffected. This wavelength is particularly suitable for telecommunication applications and renders such stimuli-responsive materials useful for transmitting encoded information [308]. In other recent experimental studies, chiral fullerenes or carbon nanotubes have been investigated by means of Raman optical activity [274, 309–312].
Chapter 6 | FIRST APPLICATION OF RVROA TO CHIRAL METAL COMPLEXES

![Molecular structure diagram]

**Figure 6.1:** Representations of the molecular structure of the Fe-helicene complex studied in this work. **Left:** Lewis structure reprinted from Ref. [276]. **Right:** B3LYP/def2-SV(P) optimized structure, C in gray, H in white, P in orange, Fe in brown.

This was the first time that RVROA was employed for examining modifications upon redox changes. In general, ROA spectroscopy, as a complementary technique to IR and VCD spectroscopy, is applied much less frequently in coordination and helicene chemistry [253, 313, 314]. However, RVROA may be a promising technique to complement experimental studies of other large chiral systems [309, 310, 312], particularly involving transition metal atoms.

Our aim is to complement the experimental findings with theoretical resonance Raman optical activity calculations in order to assign the emerging band patterns to molecular vibrations and to investigate their behaviour upon oxidation of the complex. Here, we focus on the conformer of the helicene moiety as depicted in Fig. 6.1 with (M)-helical chirality and one iron center. Furthermore, such a large chiral system puts our implementation to the test and allows us to identify possible challenges in making it a routine application in computational studies. To the best of our knowledge, this is the first theoretical RVROA study of an oxidized complex which requires spin-unrestricted complex linear response calculations.
Preparatory Steps Towards RVROA Spectra

<table>
<thead>
<tr>
<th>Mode</th>
<th>Exp.</th>
<th>( \nu_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>1355</td>
<td>1373</td>
</tr>
<tr>
<td>G</td>
<td>1594</td>
<td>1621</td>
</tr>
<tr>
<td>C-C</td>
<td>2046</td>
<td>2147</td>
</tr>
</tbody>
</table>

**Table 6.1.:** Comparison of wave numbers for the three most intense vibrations in experiment (Exp.) and returned by our calculations (\( \nu_p \)). Values are given in cm\(^{-1}\).

**Figure 6.2.:** Normal modes for the spectral bands with the highest intensity. The wave numbers corresponding to these normal modes are 1373 cm\(^{-1}\) and 1621 cm\(^{-1}\). The alkynyl stretching mode for the bond connecting the helicene to the iron center is not shown.

### 6.1 Preparatory Steps Towards RVROA Spectra

Before studying the RVROA spectra of the Fe-helicene complex, we first calculate the wave numbers corresponding to the most intense vibrations found by the experimentalists. Table 6.1 presents the calculated values compared to the experimental ones. In general, our calculations can reproduce the experimental wave numbers very well. Only for the alkynyl stretching mode, the calculations are off by almost 100 cm\(^{-1}\). However, since this vibration is isolated in the vibrational spectrum, we will still be able to analyze its RVROA intensity. Figure 6.2 shows a representation of the corresponding normal modes.
Table 6.2.: Oscillator and rotatory strengths for the Fe-helicene complex for the first ten singlet excitations. Energies are given in eV, oscillator strengths are unitless, and rotatory strengths in $1 \times 10^{-40}$ esu$^2$ cm$^2$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.9590</td>
<td>$1.56 \times 10^{-3}$</td>
<td>-13.608</td>
</tr>
<tr>
<td>2</td>
<td>1.9940</td>
<td>$7.2 \times 10^{-4}$</td>
<td>23.069</td>
</tr>
<tr>
<td>3</td>
<td>2.2764</td>
<td>$2 \times 10^{-5}$</td>
<td>-3.247</td>
</tr>
<tr>
<td>4</td>
<td>2.5014</td>
<td>$4.189 \times 10^{-2}$</td>
<td>-107.138</td>
</tr>
<tr>
<td>5</td>
<td>2.5366</td>
<td>$6.6 \times 10^{-3}$</td>
<td>10.300</td>
</tr>
<tr>
<td>6</td>
<td>2.6592</td>
<td>$1.389 \times 10^{-2}$</td>
<td>9.444</td>
</tr>
<tr>
<td>7</td>
<td>2.6835</td>
<td>$5.835 \times 10^{-2}$</td>
<td>-51.125</td>
</tr>
<tr>
<td>8</td>
<td>2.7962</td>
<td>$1.994 \times 10^{-2}$</td>
<td>-37.069</td>
</tr>
<tr>
<td>9</td>
<td>2.9196</td>
<td>$1.15 \times 10^{-3}$</td>
<td>-0.005</td>
</tr>
<tr>
<td>10</td>
<td>2.9687</td>
<td>$7.53 \times 10^{-3}$</td>
<td>3.609</td>
</tr>
</tbody>
</table>

Figure 6.3.: Experimental CD spectrum of the Fe-helicene complex (dotted green line) reprinted from the Supporting Information of Ref. [276].

As a next step, we calculate the ten lowest singlet excitations (see Table 6.2) and compare the corresponding CD spectrum to the experimental CD spectrum in Fig. 6.3. A shift of the vertical excitation energies towards lower energies can be observed. This means that we have to adjust the incident laser wavelength for calculating the RVROA spectrum to a lower energy compared to the one adopted by the experiment ($\lambda = 532$ nm) to allow for a direct comparison.

In previous studies [145], the value for the damping parameter $\Gamma$ was estimated. Yet, this only works when studying similar molecules and in the limit of short-time dynamics. In our case, we choose the damping parameter $\Gamma$ based on the experimental setup, where a Gaussian broad-
Preparatory Steps Towards RVROA Spectra

...with \( \sigma = 0.2 \) eV was applied \[276\]. The damping parameter \( \Gamma \) in our implementation corresponds to the FWHM of the Lorentzian broadening and the following relation holds between Gaussian and Lorentz broadening:

\[
\Gamma = \sigma \times 2 \times \sqrt{2 \times \ln(2)} \quad (6.1)
\]

We thus calculate for \( \Gamma \) a value of \( \Gamma = 0.47 \) eV. In atomic units, this corresponds to \( \Gamma = 0.0173 \) a.u. and we adopt this value for our RVROA calculations.

As a next step, we make sure that the \( \text{Im}(\beta) \) tensors of our response calculation behave correctly over the range of wavelength that we are interested in, i.e., between \( \lambda = 400 \) nm and \( \lambda = 600 \) nm. For computational feasibility, we restrict ourselves to the ten lowest electronic singlet excitations (see Table 6.2) and broaden them with a Lorentzian. In order to compare this to the dispersion curve of the imaginary part of the \( \beta \) tensor, we need to convert the CD intensity (units of \( 10^{40} \) esu\(^2\) cm\(^2\)) to a molar ellipticity in units of degrees \( \times \) cm\(^2\)/dmol. The necessary conversion factor can be calculated by defining the ellipticity of the polarization as:

\[
\tan \theta = \frac{E_R - E_L}{E_R + E_L} \quad (6.2)
\]

with \( E_{L,R} \) being the magnitude of the electric field vectors of the right-circularly and left-circularly polarized light, respectively. This equation can be converted by adopting Beer’s law: \[315\]

\[
[\theta] = 100 \Delta \varepsilon \left( \frac{110}{4} \right) \left( \frac{180}{\pi} \right) = 3298.2 \Delta \varepsilon \quad (6.3)
\]

Therefore, the CD intensity \( \Delta \varepsilon \) has to be multiplied by 3298.2. The optical rotation \( \beta \) which is given in atomic units must then be converted to a molar ellipticity by taking \( (1.342219 \times 10^{-4} \times v_{cm}^2 / 100.0) \times \beta \) with \( v_{cm} \) being the light frequency in units of cm\(^{-1}\). Coming from \( \omega \) in atomic units, we employ \( v_{cm} = \omega \times (27.2116 \times 8065.48049) \).
**Chapter 6**  |  FIRST APPLICATION OF RVROA TO CHIRAL METAL COMPLEXES

**Figure 6.4.** **Left:** CD spectrum (the 10 lowest electronic singlet excitation) broadened by a Lorentzian with FWHM of $\Gamma = 0.47$ eV. **Right:** Dispersion curve of $\text{Im}(\beta)$ compared to the Lorentzian of the CD spectrum. Naturally, for a perfect match of these two curves we would need to take a lot more excitations for the CD spectrum into account.
6.2 RVROA Spectra for the Neutral Fe-Helicene Complex

We now investigate the RVROA spectra for the neutral Fe-helicene complex. Even though the experimental RVROA spectrum was recorded at an incident wavelength of 532 nm, it is a priori not clear whether to adopt this wavelength for the calculations as the experimentally recorded ECD spectrum is shifted with respect to the calculated ECD spectrum, see also the discussion in Ref. [276]. Furthermore, the experimental RVROA spectrum reported in Ref. [276] did not return absolute intensities which is why a direct comparison of any calculated RVROA spectrum with the experimental counterpart is not feasible.

Therefore, we calculate the RVROA spectrum of the Fe-helicene complex over a range of incident wavelengths, starting at 580 nm and adopting an increment of 10 nm until 490 nm. In essence, we scan the VROA intensities starting from pre-resonance conditions over the first CD band. This allows us to analyse in detail how the RVROA intensities of the different normal modes behave with respect to each other and which RVROA spectrum best approximates the experimental counterpart. Naturally, we expect the RVROA intensities to vary upon changing the laser frequency of the incident light.

We first observe that all RVROA spectra corresponding to an incident wavelength of $\lambda = 580$ nm to $\lambda = 540$ nm are completely mono-signate, i.e., under full resonance or very close to it which is also shown by the experiment. Four distinct positive RVROA signatures at an incident wavelength of 580 nm can be identified, namely between 1300 cm$^{-1}$ and 1450 cm$^{-1}$, 1500 cm$^{-1}$ and 1600 cm$^{-1}$, 1600 cm$^{-1}$ and 1680 cm$^{-1}$, and a strong positive band at 2147 cm$^{-1}$, corresponding to the alkynyl stretching mode. Particularly the alkynyl stretching mode which shows the most pronounced band at 580 nm appears to be highly dependent on the incident wavelength and changes its sign from strongly positive to weakly negative at around $\lambda = 530$ nm.

Furthermore, the RVROA intensities of the normal modes corresponding to a graphene-like G-mode behave quite differently across the wavelength scan. The normal modes at 1650 cm$^{-1}$ (C–C stretching vibrations) and 1659 cm$^{-1}$ (C–C stretching vibrations coupled with C–C–H bending vibrations) show a similar RVROA intensity at 580 nm. Yet, the RVROA
Figure 6.5: Calculated RVROA spectra for the neutral Fe-helicene recorded over the incident wavelength range from 490 nm to 580 nm. The calculated spectra were broadened with a Lorentzian having a FWHM of 20 cm$^{-1}$. 
band at 1650 cm$^{-1}$ loses its intensity and even reverses its sign at 520 nm, whereas the RVROA band at 1659 cm$^{-1}$ gradually increases in intensity and still contributes to the positive RVROA signature visible at 490 nm, right next to the strong positive RVROA peak at 1621 cm$^{-1}$.

Quite interestingly, this RVROA band at 1621 cm$^{-1}$ corresponding to delocalized C–C stretching and C–C–H bending vibrations across the helicene substructure vanishes for higher incident wavelengths, particularly at 580 nm.

Beside the RVROA band pattern of the G-mode, we are also able to identify the normal modes corresponding to a graphene-like D-mode in our calculated RVROA spectra. The most prominent one of those normal modes lies at 1387 cm$^{-1}$ which has the second highest RVROA intensity at 480 nm when comparing with all RVROA intensities from 800 cm$^{-1}$ to 2400 cm$^{-1}$. Across the wavelength scan, this RVROA intensity gradually decreases, however, the strong positive RVROA signature around 1400 cm$^{-1}$ remains as the normal mode at 1409 cm$^{-1}$ gradually increases, leading to the second strongest RVROA peak at 580 nm. This normal mode is also characterized by diamond-like C–C breathing vibrations. Quite interestingly, the normal mode at 1398 cm$^{-1}$ which is also made up of a graphene-like D-mode is one of very few normal modes with a weakly negative RVROA signature.

We are also able to identify the behaviour of RVROA intensities for other normal modes of interest. For instance, the two positive peaks at 1521 cm$^{-1}$ and 1544 cm$^{-1}$ have a similar intensity at 580 nm. These modes are characterized by C–H bending vibrations across the whole molecule and C–H bending vibrations solely in the helicene substructure, respectively. However, the RVROA peak at 1544 cm$^{-1}$ loses its intensity with decreasing wavenumber until 510 nm, before it slightly increases again. By contrast, the RVROA peak at 1521 cm$^{-1}$ steadily increases its intensity and is a very prominent peak in the spectrum at 480 nm.
6.3 Towards RVROA Spectra for the Oxidized Fe-Helicene Complex

Having the RVROA spectra of the Fe-helicene complex for several different incident wavelengths, we are now able to identify which wavelength to adopt for the calculation of the RVROA spectrum of the oxidized Fe-helicene complex. By comparing the calculated RVROA intensities with respect to their experimental counterpart as given in Ref. [276] in terms of their relative intensities, we choose an incident wavelength of \( \lambda = 540 \text{ nm} \). The calculated RVROA spectrum for the oxidized Fe-helicene with this incident wavelength is shown in Fig. 6.6.

When comparing the experimental RVROA spectrum in Fig. 6.6 with our calculated RVROA spectrum, we clearly see a difference in terms of relative intensities. We expected the RVROA band of the alkynyl stretching mode to be much weaker and of comparable intensity to the D and G mode. A detailed analysis of the Hessian matrix shows that there exists several asymmetries which, in the double harmonic approximation, could be caused by a too loose structure optimization. Furthermore, we note that the coupled-perturbed KS equations did not properly converge which of course can lead to significant deviations in the numerical values for the property tensors which could become even more severe by taking numerical derivate.

Clearly, the accurate calculation of RVROA spectra for such chiral metal complex can be affected by many sources of error. It will be important to address those in order to employ the introduced methodology for production calculations. Work along those lines is currently carried out in our laboratory.
Towards RVROA Spectra for the Oxidized Fe-Helicene Complex

Figure 6.6.: **Top:** Experimental RVROA spectrum for the oxidized Fe-helicene complex (labelled as ‘M-2a’) reprinted from the Supporting Information of Ref. [276]. **Bottom:** Calculated RVROA spectra for the oxidized Fe-helicene for an incident wavelength of $\lambda = 540$ nm. The calculated spectrum was broadened with a Lorentzian having a FWHM of 20 cm$^{-1}$. 

B3LYP/SVP

$\Gamma = 0.0173$

RVROA 540 nm
Conclusions and Perspectives

Theoretical studies of transition metal complexes and their reactivity has been playing a major role in many areas of chemical research. Particularly because of the huge number of computing cores, which are now available to research groups, and the advent of elaborate statistical methods in quantum chemist [316], design studies involving a large amount of chemical structures such as high-throughput screening approaches are feasible. However, the quest for a chemical structure with a certain property can also be undertaken by inverse design approaches, one of them being the concept of gradient-driven molecule construction. The automation of this novel molecular design approach was one of the goals of this thesis.

After having introduced the theoretical foundations underlying this work, we implemented the MSD algorithm for the automated shellwise construction of molecular embedding environments to stabilize small activated (i.e., structurally distorted and/or valence unsaturated) molecular fragments of a predefined structure. While allowing for the full structural variety of the scaffolds generated by means of contiguous atom-wise additions, we may guide the search through molecular space by means of heuristic rules. These rules, however, are only required for setting up starting structures for subsequent (constrained) structure optimization and as such not an essential ingredient of the algorithm. With GdMC as scoring function, based on the nuclear fragment gradient of each generated molecular structure and the intrinsic coordination energy, the MSD framework decides on whether or not to continue the construction process. By screening the relevant chemical space in such a way, it is possible to achieve an unbiased design of chemical structures that are
able to stabilize an activated fragment such as a small inert molecule coordinated to a metal center.

Comparing with already existing concepts for molecular design, we again emphasize the key aspect which makes gradient-driven molecule construction stand out: Already at the start of the design process, we fix the design target by means of a frozen molecular fragment. That is, we encode the target geometry in the coordinates of the chemical fragment and build up a molecular scaffold around it such that the geometry gradient of the electronic energy on all nuclei of the complete molecular structure vanishes. With this at hand, we implicitly introduce a measure of comparison for all generated structures, i.e., we can clearly distinguish which structural changes favor this contrained geometry. Without the introduction of geometrical constraints, a standard full structure optimization will most often produce stationary points on the potential energy hypersurface, even though these compounds cannot be synthesized. Even worse, the desired fragment geometry could easily change in the course of an optimization such that in the end, structures with different atomic compositions for the ligand environment as well as different geometries for the fragment gradient need to be compared. Clearly, it will be difficult to determine what factors have a beneficial influence for the design task at hand.

We point out that our automated shell-wise ligand construction algorithm MSD can be applied in a very flexible fashion. First of all, the GdMC scoring function is not an integral part of our algorithm. MSD can also be applied for standard screening approaches with any suitable screening function. MSD is also very flexible with respect to the range of applications. Whereas our main focus is on design principles of transition metal catalyzed reactions, the setting described here can be applied to any problem in which a molecular substructure in a stable molecular arrangement or transition state can be used to define a chemical target problem. Lifting the GdMC scoring conditions even widens the design options. Any molecular architecture can be constructed starting from some “nucleus” structure (i.e., a single atom in the minimal version). The application of our approach will, however, be mostly chemical reaction chemistry, in which stable intermediates and transition state structures are the key players that can be well assessed based on quantum chemically accessible energies (electronic energies with and without thermal and entropic corrections). Hence, optimization, modification, or modulation
of a chemical reaction by changing the ligand or substituent environment may be cast into our framework.

In a subsequent step, we proposed a conditional design approach, in which the chemical environment of one node is constructed under the constraint that the same chemical environment is able to feature the desired properties at each of the other nodes in the reaction network. Naturally, we would like to design such a molecular scaffold for more than three generations as done in this work. Yet, already at this early stage in the design process, the combinatorial explosion made it necessary to constrain the screened molecular space. In order to limit the number of screened structures, we set thresholds for the fragment gradient $|\nabla_{\text{frag}} E_{\text{el}}|$ and the intrinsic binding energy $D_{\text{int}}^0$ based on the values calculated for a simplified version of the Schrock-type catalyst, a system which is known to fulfill the desired molecular function. Clearly, this shows the importance of introducing additional constraints as early as possible in the design process. We discussed how a local descriptor such as the Löwdin partial atomic charge on the transition metal atom could be employed in order to cut structure branches as early as possible during the design process. For future studies, we propose to enhance the selection of viable structure candidates by including a molecular fingerprint, e.g., molecular stability [317], or by strictly enforcing symmetries during the design process, something we did not do for the sake of unusual ligand environments and to show that our considerations for conditional design apply to all kinds of molecular scaffolds. In essence, the design process could be combined with an expert system that can judge whether a stable structure is also synthesizable, which may be imposed as another constraint for conditional design. This was beyond the scope of this study, but work along those lines is currently carried out in our laboratory.

We also emphasize that other design strategies, e.g., genetic algorithms [318–320], would in general not be able to carry out the introduced design principle based on local conditional descriptors. Our design concept relies on the fact that local descriptors are only modulated by structural changes in close proximity to the chemical fragment, while many other design approaches, by construction, look for structural features in the outer periphery of a ligand environment. As a direct consequence, we did not consider mechanistic effects on the scaffold center, particularly if they would happen in the outer periphery. However, it could be argued that large ligands such as the hexaisopropylterphenyl ligand in the full Schrock
complex [80, 81] could lead to steric effects and distort the molecular scaffold immediately surrounding the metal center, that is, it could lead to a change in bonding close to the activated fragment. Clearly, such effects should be the topic of later studies.

Identifying the nature and composition of a chemical compound during an experiment in a chemical laboratory poses a formidable task. A variety of methods exist which can be employed for this purpose, e.g., diffraction methods [321], crystal structure analysis [322], or spectroscopic techniques [323]. Particularly the discrimination of stereoisomers is of great importance, which can be accomplished by chiroptical spectroscopy. Here, vibrational Raman optical activity has emerged as the predominant technique besides other chiroptical techniques, e.g., electronic and vibrational circular dichroism and optical rotation. In the second part of this thesis, we introduced a new implementation for the calculation of resonance-enhanced Raman optical activity spectra. The VROA and RVROA spectra for hydrogen peroxide, methyloxirane, (S)-naproxen, and (S)-ibuprofen were presented and the experimentally observed band patterns could be readily interpreted. In accord with experiment, we were able to predict the resonance-enhanced behavior of the VROA spectrum of (S)-naproxen at an incident wavelength of 514.5 nm, with mostly monosignate character, while at the same time, (S)-ibuprofen showed a typical off-resonance spectrum at the same incident wavelength. All spectra were calculated in the short-time approximation to the time-dependent theory of Raman scattering which is identical to the standard theory introduced by Placzek in the off-resonance case, while it becomes a Placzek-like polarizability theory under resonance conditions.

Even though more complicated and involved molecular structures are now accessible by means of VROA spectroscopy, for example ribonucleic acids which contain a wealth of structural information [324, 325], the missing direct relation between a structure and its spectral pattern and in particular the sensitivity of a VROA spectrum very often lead to a preference for spectroscopic techniques in practice. For example, a single positive VROA signal characteristic for a certain molecular feature [236, 237] can easily be cancelled out by a closely lying negative peak. Furthermore, a biomolecule is in general made up of many secondary structural features and not just a single α-helix or β-sheet. This in turn leads to an
overcrowded VROA spectrum with lots of different band patterns making a detailed analysis of such a spectrum even more complicated.

However, resonance-enhanced VROA could become a new technique for the identification of transition metal complexes [261]. Since not too many theoretical studies exist on this topic, it is certainly a worthwhile path to explore further. To this end, we calculated the RVROA spectra for a chiral transition metal complex, that is, an ethynyl-helicene mono-FeII complex grafted with one electroactive [Fe(κ^2-dppe)(η^5-C_5Me_3)] fragment. During this process, we identified and discussed several issues which still need to be elaborated on before such an implementation can be employed in a production environment.

We think that we were able to show the importance of the introduced exploratory and spectroscopic tools in order to address the future challenges of this field. Clearly, the combination with other quantum chemical concepts such as an error estimation of chemical reaction energies [326], the automatic exploration of chemical reaction networks [7, 327], or the haptic exploration of potential energy surfaces [328–333] into one single virtual reactivity laboratory will be an invaluable tool for any chemist. Work along those lines is currently carried out in our laboratory.
Computational Methodology

In the following, we summarize the computational methodology employed for each chapter.

A.1 Molecular Scaffold Design by Gradient-driven Molecule Construction

A.1.1 Generating a Molecular Environment by Frozen Density Embedding

All FDE calculations were carried out with the density functional theory programs provided by the Amsterdam Density Functional software suite (ADF)[334] and the PyADF framework [335]. Freeze-and-thaw cycles were not performed. The PW91 [336] functional and the TZP-basis set [337] were employed, both of which proved to be a good choice for FDE calculations involving hydrogen bonds [338]. For the iterative potential optimization, we employed the Nelder–Mead simplex algorithm [183] where in each step the updated potential was fed into a modified version of TURBOMOLE [339] as an external potential. The PBE functional [340, 341] and the def-TZVP basis [342] set were employed here.
A.1.2 Constructing the Chemical Environment by Atom-wise Additions

We implemented the algorithm described in the Section 3.2 as a C++ program to fully automate the construction of an embedding environment. BP86 raw-data calculations [340, 341] were carried out with the density functional theory programs provided by the TURBOMOLE software suite [343] (version 6.4.0) and the def2-TZVP [344] basis set. The resolution-of-the-identity density-fitting technique was invoked in all cases. A relativistic effective core potential was chosen for the ruthenium atom [345].

For each optimization, the spin quantum number $S$ was determined as $S=M_S$ from the lowest number of unpaired electrons that may be distributed over degenerate frontier orbitals (i.e., a singlet state was chosen for a non-degenerate highest occupied molecular orbital). Hence, higher spin states that might be lower in energy were not considered so that the automated design occasionally might produce a structure in a spin state that is an excited state. For our Ru complex, this should not be a pressing issue as heavy-element complexes often prefer ground states with lowest spin quantum numbers, but it must be considered for other systems and the program can be easily extended to account for ligand construction in different spin states.

The activated fragment was positionally constrained by means of four frozen internal coordinates corresponding to an activated CO$_2$ fragment. For this structural activation we chose a Ru-C distance of 205.9 pm, a C-O distance of 120.9 pm, a C-$\eta$(O) distance of 126.8 pm, and an O-C-O angle of 139.73°, which were obtained from a quantum chemical structure optimization of the full complex with a different density functional and basis set. The idea is to provide a reasonable activated fragment that will not be exactly the same as one would obtain for the DFT setting used through this study because, in general, one will not know the target structure and needs to make some reasonable choice for the structure of the activated fragment.

If the definition of internal coordinates turned out to be unfeasible (e.g., due to linear dependencies), the algorithm automatically switched to a constrained optimization in Cartesian coordinates. I.e., the activated fragment was then positionally constrained by fixing the position of all fragment atoms in Cartesian coordinate space. We considered a structure
optimization converged when the norm of the electronic energy gradient with respect to the nuclear coordinates that were left free to relax was below $10^{-4}$ hartree/bohr. A subsequent single-point calculation without any constraints on the internal coordinates returned the nuclear gradient of the electronic energy on the activated fragment. Here, we considered a single-point calculation converged once the total electronic-energy difference between two iteration steps was smaller than $10^{-6}$ hartree.

The intrinsic coordination energy $D^\text{int}_e$ of the (bent) CO$_2$ ligand to the central metal ion including the embedding environment was calculated as the electronic energy difference of the frozen (separated) structures. We also determined the electronic coordination energy $D_e$ for the fully relaxed structures.

If a self-consistent-field calculation did not converge after 1’000 iterations, the structure was discarded and not considered for subsequent generations. The visualization of all figures concerning the position of a structure in the space of fragment gradient and intrinsic binding energy was automated in the programming language PYTHON. When applicable, the distribution of structures was visualized by the standard deviational ellipse [211].

A.2 Conditional Molecular Scaffold Design

All density functional theory (DFT) calculations were carried out with the BP86 density functional [340, 341] provided by the TURBOMOLE software suite [343] (version 6.4.0) and the def2-TZVP [344] basis set combined with the resolution-of-the-identity density-fitting technique. An effective core potential was chosen only for the molybdenum atom [345].

For each optimization, we chose the spin quantum number $S$ as $S = M_S$ by distributing the lowest possible number of unpaired electrons over degenerate frontier orbitals which always returned a singlet state for a non-degenerate highest occupied molecular orbital. Therefore, we did not consider higher spin states possibly lower in energy, however, we expect the Mo-complexes investigated here to prefer ground states with the lowest spin quantum number. The MoN$_2$ and MoNH$_3$ fragments were positionally constrained by means of frozen internal coordinates corresponding to an activated N$_2$ or bound NH$_3$ fragment, respectively. Here, the choice of fixed Mo–N and N–N bond lengths is essential since
there are two obvious choices for the Mo-N bond length. Either one chooses a bond length known for similar dinitrogen binding complexes, in our case from the original Schrock catalyst (198.4 pm, see Ref. [229]), or one can let this bond length relax to its equilibrium value in a given model complex. Furthermore, one can set the N-N bond length to a value indicating a certain desirable degree of bond activation, such as in the original Schrock complex (114.2 pm, see Ref. [229]). Here, the N–N distance represents an activated molecular dinitrogen ligand (as a comparison, the N–N distance for isolated N₂ is calculated to be 110.4 pm in Ref. [229]). The same consideration applies to the Mo-N and N-H distances for the coordination of NH₃ to Mo. We adopt a Mo-N distance of 228.8 pm and a N-H distance of 103.2 pm, respectively, as calculated in Ref. [229].

It should be noted that the number of possible structures can be reduced during the design process by restricting the oxidation state of the transition metal center to a certain formal charge, e.g., +III as in the original Schrock catalyst. In general, the formal oxidation state of a transition metal center is determined by the total charge of the complex and the heterolytically cleaved metal-ligand bonds. Since we are interested in general concepts during our design process, we do not intend to constrain the formal oxidation state of Mo to a certain formal charge, but allow for the construction of catalysts with all experimentally observed oxidation states of the molybdenum atom.

The design process was carried out by employing our fully automated molecular scaffold design program GdMC-MSD. We restricted the number of child structures for each parent structure to 1’000 for the second generation. In case more than 1000 child structures could be formed out of one parent structure, we selected 1’000 structures out of the complete set of child structures by means of simple random sampling. For the third generation, the same procedure was applied, but with a maximum value of 100 structures. This procedure has two advantages: First, it allowed computational feasibility of the design study. Second, by sampling over all possible structures, we could make sure to include ligand environments that represent the complete set of possible child structures in a fairly accurate way. Furthermore, we applied certain chemical heuristics, e.g. the group PH₃ is only symmetrically extended to yield P(CH₃)₃.

We considered a structure optimization converged as soon as the norm of the electronic-energy gradient with respect to the nuclear coordinates
was below $10^{-4}$ Hartree/Bohr. A subsequent single-point calculation returned the geometry gradient of the electronic energy on the chemical fragment. Here, we considered a single-point calculation converged once the total electronic-energy difference between two iteration steps was smaller than $10^{-8}$ Hartree. The intrinsic coordination energy $D_e^{\text{int}}$ of the N$_2$ and NH$_3$ fragments to its embedding environment was determined by calculating the electronic energy $E_{\text{frag},\text{frozen}}$ of the frozen N$_2$ and NH$_3$ fragments and of the constraint-optimized structure without the N$_2$ and NH$_3$ fragments ($E_{\text{lig},\text{frozen}}$), respectively, as single points and subtracting them from the electronic energy $E_{\text{frag}−\text{lig},\text{relaxed}}$ of the constraint-optimized structure as follows:

$$D_e^{\text{int}} = E_{\text{frag}−\text{lig},\text{relaxed}} - E_{\text{frag},\text{frozen}} - E_{\text{lig},\text{frozen}}.$$  \hspace{1cm} \text{(A.1)}$$

All atomic replacements and the introduction of structural features within the GdMC-MSD framework are based on idealized coordination geometries. In addition, we applied chemical intuition during the design process, e.g., there must not appear a chain of four oxygen atoms at any time. For occupying the binding sites in each generation, we define our atom set as \{H, C, N, O, P\} or a subset thereof. We saturated any remaining unsaturated valencies by hydrogen capping atoms. Inspired by the Schrock-type catalyst, benzene derivatives were employed to functionalize the scaffold structures instead of additionally binding single atoms from the third generation onwards. Once a set of structures was created, they were automatically subjected to a constrained optimization, which froze the fragment to be stabilized, but relaxed the constructed scaffold structure. The geometry gradients of the electronic energy for the chemical fragment as well as the (intrinsic) coordination energies were then taken for the scoring function.

### A.3 Stereoisomer Discrimination by Chiroptical Vibrational Scattering Spectroscopy

We implemented the theory described in Section 5 in our MoViPAC program package [288] with which all calculations of the VROA and RVROA intensities were performed. Structure optimizations and response calculations were performed at the KS DFT level with the BLYP density
Appendix A | COMPUTATIONAL METHODOLOGY

functional and the def2-TZVP Gaussian-type basis set, employing a locally modified version of the NWChem program [346]. The modifications of NWChem entail, first, the interface with MoViPac in order to provide the calculated complex response tensors of Eqs. (2.74) – (2.76). These tensors were calculated with KS DFT using a complex response module implemented by one of us [347, 348]. The NWChem KS response module supports non-hybrid, global hybrid, and range-separated hybrid functionals, and the $G'$ tensor can be calculated with frequency-dependent gauge-including atomic orbital (GIAO) basis functions [140] in order to obtain origin-invariant (R)VROA spectra [145], or with the standard dipole-length gauge. The second modification of NWChem is a new implementation for $G'$ based on Eq. (57) in Ref. [140]; previously $\beta(\omega) = -G'/\omega$ was determined by a procedure [347] that required two response calculations, one dynamic and one with $\omega = 0$, and converted to $G'$.

After test calculations showed that the use of a GIAO basis is not vital as long as the molecules are not very large, as is the case here, and centered at the coordinate origin, the dipole-length gauge was chosen as it entails computational savings. The BLYP non-hybrid functional was chosen because it gave accurate results in RVROA benchmark calculations in the past [145]. Single-point calculations were considered converged when the total electronic-energy difference between two iteration steps was less than $10^{-10}$ Hartree, unless stated otherwise. For the structure optimizations, the default thresholds for the maximum and root mean square gradient ($4.5 \times 10^{-4}$ a.u. and $3.0 \times 10^{-4}$ a.u.) as well as the maximum and root mean square of the Cartesian step ($1.8 \times 10^{-3}$ a.u. and $1.2 \times 10^{-3}$ a.u.) were employed.

Values of the damping parameter $\Gamma$ in the response calculation have to be obtained by fitting absorption data for the molecule. However, in the limit of short-time dynamics and when studying similar molecules, an estimated value of $\Gamma = 0.0037$ a.u. worked well in previous studies [145] and we applied it in this work as well. The polarizability derivatives were calculated by adopting a numerical three-point differentiation with respect to Cartesian displacements and a step size of $s = 0.01$ a.u. [280].

The line spectra resulting from this methodology were broadened by means of a convolution with a Lorentzian line shape featuring a full width at half-maximum (FWHM) of 20 cm$^{-1}$. For the line broadening and graphical representation of all spectra, we employed the program math-
EMATICA (version 9.0.1) [349]. Atomic displacements for the analyzed vibrational normal modes were visualized with Jmol [350].

### A.4 First Application of RVROA to Chiral Metal Complexes

Structure optimizations and response calculations were performed at the KS DFT level, employing a locally modified version of the NWChem program [346] providing complex response tensors [140, 145, 347, 348]. A convergence threshold of $1.0 \times 10^{-6}$ a.u. was chosen for the iterative solution of the linear response equations. Furthermore, we adopted a value of $\Gamma = 0.0173$ a.u. for the damping parameter $\Gamma$ in the response calculations which corresponds to a Gaussian broadening with $\sigma = 0.2$ eV employed by the experimentalists [276].

A comparison of experimental and simulated ECD spectra for the systems studied (see Supporting Information of Ref. [276]) showed that most calculated results are clearly deficient except those obtained with B3LYP which is why we also employ this functional for this study, together with the def2-SV(P) Gaussian-type basis set.

All RVROA backscattering intensities were calculated by employing the MOVIPAC software suite [288]. This allows for a massively parallel calculation of RVROA spectra based on numerical derivatives to evaluate the spectroscopic intensities as first derivates of the respective property tensors. The polarizability derivatives were calculated by adopting a numerical three-point differentiation with respect to Cartesian displacements and a step size of $s = 0.01$ a.u. [280]. The dipole-length gauge was chosen in order to allow for computational savings.

Single-point calculations were considered converged when the total electronic-energy difference between two iteration steps was less than $10^{-10}$ Hartree, unless stated otherwise. For the structure optimizations, the default values for the maximum and root mean square gradient ($4.5 \times 10^{-4}$ a.u. and $3.0 \times 10^{-4}$ a.u.) as well as the maximum and root mean square of the Cartesian step ($1.8 \times 10^{-3}$ a.u. and $1.2 \times 10^{-3}$ a.u.) were employed.

The line spectra resulting from this methodology were broadened by means of a convolution with a Lorentzian line shape featuring a full width at half-maximum (FWHM) of 20 cm$^{-1}$. For the line broadening and
Appendix A  | COMPUTATIONAL METHODOLOGY

A graphical representation of all spectra, we employed the program Mathematica (version 9.0.1) [349]. Atomic displacements for the analyzed vibrational normal modes were visualized with Jmol [350].
The following publications are included in parts or in an extended version in this thesis:


The work presented in chapter 4 has been carried out in collaboration with Prof. Dr. Markus Reiher, ETH Zurich, and is still to be published. The work presented in chapter 6 has been carried out in collaboration with Prof. Dr. Markus Reiher, ETH Zurich, and Prof. Dr. Jochen Autschbach, University of Buffalo, and is still to be published.

Additional publications which are not part of this thesis:

Acknowledgments

This thesis would not have been possible without the constant support and encouragement by Prof. Dr. Markus Reiher. His enthusiasm for theoretical chemistry has inspired me throughout my time in his group. He always had an open ear for any kind of question and was a great supervisor. It was a privilege to work with him for which I am extremely grateful.

Also, I had the privilege to have worked with a lot of wonderful people over the last few years. In particular, I would like to thank Dr. Moritz Haag for being a great supervisor during my Master Thesis and for introducing me to the Reiher group. For all countless discussions about quantum chemistry, science, and life in general, I say thanks to Dr. Martin Stiebritz and Dr. Thomas Weymuth.

A warm thanks to all the other past and current group members which I had the privilege to work with, in alphabetical order: Dr. Maike Bergler, Dr. Katharina Boguslawski, Dr. Marta Bruska, Dr. Leon Freitag, Tamara Husch, Dr. Erik Donovan Hedegard, Dr. Sebastian Keller, Dr. Stefan Knecht, Dr. Arseny Kovyrshin, PD Dr. Hans-Peter Lüthi, Dr. Yingjin Ma, Dr. Halua Pinto de Magalhães, Adrian Mühlbach, Andrea Muolo, Jonny Proppe, Dr. Oliver Sala, Gregor Simm, Dr. Benjamin Simmen, Dr. Pawel Tecmer.

Furthermore, I would like to thank Romy Isenegger for taking care of all administrative issues. For keeping the computer system running, I would like to thank Dr. Moritz Haag, Dr. Thomas Weymuth and Dr. Sebastian Keller.

A special thanks goes to Pascal de Buren, Patrick Seewald, and Jan-Grimo Sobez, who have been great students and who contributed a lot to this work. They all did an excellent job and I had a great time working together with them.

My deepest appreciation goes to my parents, Manfred and Vera, and my brother Martin. Without their constant support and belief in me, I would have not endured so long. They have always been here for me.
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


Colophon

This document was typeset in \TeX using the typographical look-and-feel classicthesis. The bibliography is typeset using biblatex.