Theoretical studies for the development of a better understanding of cofactor-structure and mechanistic properties of photoreceptors and metallo-enzymes with quantum chemical and molecular dynamical approaches

vorgelegt von
Master of Science-Chemiker
Dennis Heinz Belger

an der Fakultät II – Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. rer. nat. Arne Thomas
Gutachterin: Prof. Dr. rer. nat. Maria Andrea Mroginski
Gutachterin: Prof. Dr. rer. nat. Silke Leimkühler

Tag der wissenschaftlichen Aussprache: 29.11.2019

Berlin 2020

Table of Contents

IV. Molecular dynamics and QM/MM based computations for phytochrome Agp2

V. QM calculations and subsequent Raman-spectra computations *via* NMA for *Rc*FDH

List of Abbreviations.

Phytochrome	Phy
Cph1	*Cyanobacterial phytochrome 1*
PCB	Phycocyanobilin
Agp2	*Agrobacterium tumefaciens phytochrome 2*
BV	Biliverdine
Rc	*Rhodobacter capsulatus*
FDH	Formate dehydrogenase
p-water	Pyrrole-water
CH_3	Methyl group
RR	Resonance Raman
eq.	Equation
MM	Molecular mechanics
MD	Molecular dynamics
polff	Polarized force field
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RGYR	Radius of gyration
HF	Hatree Fock
DFT	Density functional theory
QC	Quadratically convergent
SCF	Self consistent field
PBC	Periodic boundary conditions
SBC	Stochastic boundary conditions
PME	Particle Mesh Ewald approximations
QM	Quantum mechanics
PCM	Polarized continuum model
MM-MD	MD-simulation with classical molecular mechanics Force field
polff-MD	MD-simulation with polarized force field
His	Histidine
Asp	Aspartate
Cys	Cysteine
Ser	Serine
Pro	Proline
Arg	Arginine
Tyr	Tyrosine
Lys	Lysine
OH^-	Hydroxyl
CO	Carbonyl
CH_2	Methylene
NH-ip	N-H in-plane rocking
HSE	Nε-protonation of histidine
HSD	Nδ-protonation of histidine
HSP	Double protonation of histidine
ed, ee, de, etc.	Corresponding protonation states of N of His248 and His278 (f.e. ed means His248 protonated corresponding to HSE and His278 protonated corresponding to HSD)

PSC (B/C)	Propionic side chain (ring B or C) of biliverdine
bimme	Biliverdine monomethylester
BVMB	Biliverdine monomethylester methylated at PSC(B)
BVMC	Biliverdine monomethylester methylated at PSC(C)
BiMET	Doubly methylated biliverdine at PSC(B) and PSC(C)
RR	Resonance Raman
R	Raman
IR	Infrared
Subtr.	Subtraction
Moco	Moybdenum cofactor
XAS	X-ray absorption spectroscopy
EXAFS	Extended X-Ray absorption fine structure
EPR	Electron paramagnetic resonance
wo	Without
GDP	Guanine diphosphate
small Moco	Moco without GDP units
bigger Moco	Moco without guanine parts, but with phosphates and ribose
S	Sulfido
O	Oxo
OH	Hydroxo
SH	Thiol
PO_4	Phosphate

Abstract

Diversity of life is not only defined by the number of genes in an organism, but by the proteins and subsequent modifications of these. The different proteins and different functionality are the key for generating life as it is known. Proteins are important for structural stability of cellular systems, as well as for mediating all kinds of processes (like chemical reactions and sensing of external stimuli). Understanding of the structure and functionality of proteins on an atomic level is of utter importance for developing pharmaceuticals, artificial devices and their applications. In this work different protein systems (such as photosensors and enzymes) are investigated with theoretical methods which in the past have proven very useful for determination of structure and function of large biomolecules. The results from theory can then verify, complete or predict experimental results, such as obtained by Raman spectroscopy and X-ray analysis.

Phytochromes are biological red/far-red light sensing proteins occurring in plants, as well as in bacteria and fungi. In this work two phytochromes, namely Cyanobacterial phytochrome Cph1 and Agrobacterium tumefaciens phytochrome Agp2, are investigated. In Part III of this work, the development and its application of a new polarized force field for the chromophore-binding domain (CBD) of Cph1 is shown. In the case of Cph1, the stability of the water-network inside the binding domain of the chromophore has been addressed with a polarized force field which has been developed from Mulliken charges derived by quantum chemical computations of the cofactor and its immediate environment. MD-simulations with this force field revealed an increased stability of the pyrrole-water. This stabilization effect can be seen in computed Raman-spectra. Hybrid quantum mechanic/molecular mechanic (QM/MM) approaches have been used for subsequent Raman-spectra calculations. The implementation of polarization effects led to a better agreement of computed spectra with experiments, especially in terms of the NH-in-plane rockings (more pronounced bands) which are due to less fluctuating water-molecules. Thus, this method improves existing MD-protocols by exchanging conventional point charges of atoms with quantum chemically derived partial charges leading to a stabilization of the water network and can be used for more accurate spectra-computations.

Part IV illustrates the investigation of the Agp2 biliverdine-binding domain in its parent Pfr-state. QM/MM methods have been used for determination of the correct protonation state of the two conserved histidine residues His248 and His278 in near vicinity of the chromophore. Comparison of different protonation models of His248 and His278 with the corresponding X-ray structure. revealed that the former histidine is protonated at the N_{ε}, while the latter carries a proton on N_{δ}. Furthermore, QM/MM- and MD-techniques have been applied to different variants of Agp2, in which BV carries a methyl group on either the propionic side-chain of rings B (BVMB) or C (BVMC) or both (BiMET). Comparison to the crystallographic structure with native-cofactor inside its binding-domain, revealed a high flexibility of Arg211 to the introduction of the methyl function at propionate of ring B. It induced an opening of the chromophore binding domain not be observed in the wild-type. In contrast to this, the variant BVMC showed big deviations of a highly unstable His278 from its original position and less tolerance to the additional methyl group. This led to the hypothesis that BVMB is photoactive, because of its ability to produce a stable CBD after initial rearrangements. BVMC was postulated to be photoinactive, due to the highly unstable CBD. This is in agreement to experimental findings proposing BVMB to undergo a complete photoconversion upon light-irradiation, wile BVMC remains photoinactive.

*Studies of formate dehydrogenase of Rhodobacter capsulatus (Rc*FDH*) are shown in Part V of this work. Rc*FDH *is an enzyme which catalyzes the reversible oxidation of formate to carbon dioxide upon reduction of NAD^+. It carries a Mo-containing cofactor, called Moco, common in FDHs of*

*different organisms. Since no crystal structure is available for Rc*FDH *so far, the complete coordination sphere of the Mo has not been determined conclusively, yet. Here, a combination of quantum mechanical (QM) geometry optimizations and computations of Raman-spectra has been done, using a homology model of RcFDH. Different models of a smaller Moco (without GDP-units), with different ligation spheres have been built, including the native Cys386 in direct vicinity of the cofactor. Subsequent comparison to experimental resonance Raman findings revealed, the ligands of Mo: Mo exists in at least two different oxidations states +VI and +IV. Both are coordinated by two dithiolene moieties. MoVI is further ligated to a native Cys386 residue and either a sulfido-ligand or an oxo-ligand. The latter leads to an inactive species which has been observed in experimental EXFAS studies. The MoIV-species lacks (at least temporarily) the Cys-ligation and is pentacoordinated by a sulfido-ligand (or temporarily thiol-group) and two dihtiolenes. Additional models with different ligands at the metal-site for both oxidation states, as well as bond dissociation enthalpy computations of different ligand-metal bond systems gave deep insight into the catalytic cycle: Formate does not bind directly to MoVI. In contrast to that, Cys386 stays coordinated to MoVI, while in MoIV Cys leaves the metal. This is in accordance with a mechanistic proposal that formate stays unbound near to the metal-center during the reaction cycle. Possible formal hydride-transfer from formate is mediated by the sulfido-ligand leading to the reduction of molybdenum to MoIV and temporarily forming a SH-ligand at MoIV which further translocates a proton to Cys386 and afterwards deeper into the protein. Subsequently, Cys386 is returning to the metal-site and two electrons are released from the metal and the cycle can start anew (while* CO_2 *is released from the active-site). The on-and-off-going of Cys386 may be mediated by the formation of a disulfide-bridge with the sulfido-ligand. Also, the possibility of an intermediate MoV-species has been investigated and possible participation in the mechanism discussed. Finally, the complete coordination spheres of both MoVI, MoV and MoIV in the Moco of Rc*FDH *are shown, with possible influence on the reaction cycle. The here presented catalytic mechanism includes important characteristics of recently discussed ones (like proton-abstraction mediated by the Cys-group and temporal formation of a SH-group and formal hydride-transfer and only indirect formate-Moco interaction), as well as new feautres and thus represents an improved and extended version.*

Zusammenfassung

Die Vielfalt des Lebens wird nicht allein durch die Anzahl der Gene in einem Organismus bestimmt, sondern vor allem durch die unterschiedlichen Proteine und deren Modifikationen. Unterschiedliche Proteine und deren Eigenschaften sind entscheidend für die Erschaffung von Leben, so wie es bekannt ist. Proteine sind wichtig für die Stabilität von zellulären Systemen, genauso wie für die Durchführung unterschiedlichster Prozesse (wie zum Beispiel chemische Reaktionen und das Erkennen von externen Reizen). Um Medikamente und artifizielle Proteine herzustellen, sowie für deren Anwendung ist es von unerlässlicher Bedeutung, die Struktur, sowie die Funktionalitäten von Proteinen auf einem atomaren Level zu kennen. In dieser Arbeit sind verschiedene Proteinsysteme (wie zum Beispiel Fotosensoren und Enzyme) mit Hilfe von theoretischen Methoden untersucht worden. In der Vergangenheit hat sich der Einsatz von theoretischen Berechnungsmodellen als sehr hilfreich bei der Bestimmung von Struktur und Funktion großer Biomoleküle erwiesen. Die Ergebnisse von diesen Berechnungen können dann genutzt werden um experimentelle Studien zu vervollständigen, zu bestätigen oder sogar vorherzusagen, wie es zum Beispiel bei Raman-Spektroskopie und der Röntgenstrukturanalyse in dieser Arbeit getan wurde.

Phytochrome sind biologische lichtempfindliche Proteine, die im roten Spektralbereich absorbieren. In dieser Arbeit werden zwei verschiedene Phytochrome untersucht, jenes von Cyanobakterium (Cph1) und das von Agrobacterium tumefaciens (Agp2). Im Teil 3 dieser Arbeit wird die Entwicklung eines neuartigen polarisierten Kraftfeldes für die Chromophor-Bindungs-Domäne (CBD) von Cph1 dargestellt. Mulliken-Ladungen des Chromophores und seiner direkten Umgebung wurden mittels quantenmechanischer (QM) Rechnungen ermittelt und zur Entwicklung eines polarisierten Kraftfeldes herangezogen. Dieses wurde in einer MD-Simulation genutzt um die Stabilität des Wassernetzwerks innerhalb der CBD zu adressieren und zu erhöhen. Im Ergebnis erhöhte sich die Stabilität des Pyrolwassers deutlich, was auch in nachfolgend berechneten Ramanspektren gesehen werden konnte. Eine Methode bestehend zum Teil aus quantenmechanischen Ansätzen und zum anderen Teil aus klassisch molekülmechanischen (QM/MM) wurde genutzt um Ramanspektren für Cph1 zu berechnen (basierend auf der Simulation mit polarisiertem Kraftfeld). Die hiermit berechneten Spektren zeigen eine deutlich bessere Übereinstimmung mit experimentellen Daten und weisen eine deutlich schmalere NH-in-plane Schwingungsbande auf. Dies ist auf die Stabilität des Pyrrolwassers auf seiner nativen Position zwischen den Pyrrolringen zurückzuführen. Die hier vorgestellte Methode konventionelle Punktladungen in MM-Kraftfeldern gegen mit QM-Methoden berechnete Partialladungen auszutauschen, stellt demnach eine Verbesserung gängiger Simulationsmethodiken von Biomolekülen dar und ermöglicht eine realistischere Berechnung von Schwingungsspektren.

In Teil 4 dieser Arbeit wird die Untersuchung der CBD von Agp2 in seinem nativen Pfr-Zustand diskutiert. QM/MM Techniken sind genutzt worden um den Protonierungszustand der beiden konservierten Histidine His248 und His278 in unmittelbarer Nähe des Chromophores Biliverdin (BV) zu bestimmen. Verschiedene QM/MM-Modelle, die sich durch mögliche Protonierungen der Stickstoffe der beiden Histidine unterscheiden, wurden mit der Kristallstruktur von Agp2 verglichen. Hierbei wurde ermittelt, dass His248 am N_ε protoniert ist, während His278 sein Proton am N_δ trägt, da das dazugehörige Modell die geringsten strukturellen Abweichungen von der Kristallstruktur aufwies. Durch die Identifizierung des korrekten Protonierungszustandes in der CBD von Agp2 ist es nun möglich weiterführende theoretische Untersuchungen anzustellen, was der Interpretation von experimentellen Befunden zu Gute kommt. So wurden drei verschiedene Modelle erstellt, welche sich in der Position einer unnatürlichen Methylgruppe an den Propionatseitenketten des Kofaktors unterschieden: an ring B (BVMB), ring C (BVMC) oder an

beiden (BiMET). Diese wiederum wurden mit Hilfe von MD- und QM/MM-Methoden untersucht und mit der Kristallstruktur mit unmethyliertem Kofaktor verglichen. Dabei zeigte sich, dass Arg211 eine besonders hohe Flexibilität aufweist, in Bezug auf die eingeführte Methylgruppe. Letztgenannte Gruppe induziert eine Öffnung der CBD unter Einbezug von Arg211 (ähnlich wie bei einem Fass, an welchem der Deckel angehoben wird), welche in dieser Art nicht im nativen Protein simuliert werden konnte. Im Gegensatz dazu, offenbarte die Simulation von BVMC ein sehr instabiles His278, welches starke Abweichungen zu seiner Ursprungsposition zeigte. Weite Teile der CBD in BVMC zeigten deutliche Instabilitäten während der MD- und QM/MM-Berechnungen. Dies führte zu der Hypothese, dass BVMB fotoaktiv sei, da sich hier die CBD der eingeführten Methylgruppe anpassen konnte und am Ende der MD stabil blieb. Im Gegensatz dazu sei BVMC auf Grund der Instabilität in der CBD nicht in der Lage eine komplette Fotokonversion durchzulaufen. Dies ist in Übereinstimmung unveröffentlichten Experimenten.

*In Teil 5 dieser Arbeit wird das Enzym Formiatdehydrogenase von Rhodobacter capsulatus (Rc*FDH*) eingehend untersucht. Dieses Enzym ist verantwortlich für die reversible Oxidation von Formiat zu Kohlenstoffdioxid. Hierbei wird NAD^+ zu NADH reduziert. Der Kofaktor, welcher in allen bekannten FDHs zu finden ist, der sogenannte Moco, trägt ein Molybdän in seinem Zentrum. Die genaue Koordinationssphäre des Mo's im Moco ist noch nicht endgültig bekannt, da es zum jetzigen Zeitpunkt keine Kristallstruktur von RcFDH gibt. Deshalb wurde ein Homologiemodell von RcFDH genutzt um verschiedene Modelle eines kleineren Mocos zu generieren. Mittels QM-Methoden berechnete Raman-spektren wurden hinsichtlich ihrer Vergleichbarkeit mit experimentellen Resonanz-Raman-Spektren untersucht, um so die korrekte Koordinationssphäre des Metalls zu entschlüsseln. Hierbei kann das Metall mindestens in zwei verschiedenen Oxidationsstufen vorliegen, entweder oxidiert als +VI oder reduziert als +IV. Beide Metallformen sind von zwei Dithioleneinheiten koordiniert. MoVI ist zudem vom Cys386 direkt koordiniert und trägt zusätzlich entweder einen Sulfido- oder Oxo-Liganden und ist somit hexakoordiniert. Der letztgenannte Ligand wurde in EXFAS-Messungen in einer inaktiven Form des Enzyms beobachtet. MoIV hingegen trägt (zumindest teilweise) keinen Cys-Liganden und ist pentakoordiniert. Das Metall ist von zwei Dithiolenen und einer Sulfido-Gruppe (evtl. auch temprär von einer Thiol-Gruppe) koordiniert. Mit Hilfe dieser Erkenntnisse wurden weitere Modelle erstellt, die sich hinsichtlich zusätzlicher Liganden in beiden Oxidationsstufen unterschieden. Zusammen mit der Berechnung von Bindungsbrechungsenthalpien verschiedener Metall-Liganden-Bindungen lieferte dies neue Erkenntnisse hinsichtlich eines möglichen Reaktionsmechanismus: MoVI hat eine stärkere Bindung zu Cys386, als zu Formiat, weshalb Cys386 nicht als Ligationspartner des Metalls von Formiat ausgetauscht wird. Dies ist in Übereinstimmung mit bereits veröffentlichten Daten. Weiter konnte gezeigt werden, dass ein Hydridtransfer des Formiatprotons mittels des Sulfido-Liganden unter Ausbildung einer temporären SH-Gruppe am Moco möglich ist. Dadurch wird MoVI zu MoIV reduziert. Das Proton wird dann sehr schnell weiter zum negativ geladenen Cys386 geleitet, welches nicht stabil an MoIV bindet, von wo es tiefer in das Protein transportiert wird. Auch Formiat (und evtl. CO_2) bindet nicht stabil an das reduzierte Metallzentrum, was zu einer pentakoordinierten Form des MoIV führt. Nach dem Verlassen von CO_2 koordiniert Cys386 wieder an das Metall, welches zwei Elektronen abgibt (MoIV → MoVI) und der Zyklus kann von vorne beginnen. Die Abspaltung und Bindung von Cys386 vom bzw. zum Metall könnte dabei mittels der Ausbildung einer Disulfidbrücke mit dem Sulfido-Liganden vonstatten gehen. Ebenfalls ist es möglich, dass die Elektronen Schritt für Schritt übertragen werden und somit eine MoV-Zwischenstufe präsent ist. Diese Möglichkeit wurde getestet und ein möglicher Anteil am Mechanismus diskutiert. Zusammenfassend lässt sich sagen, dass der hier vorgestellte Mechanismus Eigenschaften der beiden vorherrschenden Katalysemodelle (Protontransfer mittels Cys386, SH-Ligand-Ausbildung, Hydridtransfer, keine Ligation von Formiat an Moco) zusammenführt und noch um neue wichtige Charakteristika erweitert und somit verbessert.*

Part I.

Introduction

Chapter 1

Phytochrome Photoreceptors

1.1. General protein structure motives (Cph1 and Agp2)

Phytochromes are biological light-receptors. They are responsible for determination of red light and management of different physiological functions in both higher plants [1] and bacteria, fungi and algae [2],[3]. In the former they are responsible e.g. for avoidance of shade and orientation towards a light-source, the development of seeds and flowers and leave-formation. In organisms other than plants, like bacteria for example, they mediate e.g. orientation in the light-deficient soil. These photoreceptors interconvert reversibly between a red-light sensitive Pr state and a far-red-light sensitive Pfr state, where one state is inactive (but thermally stable) and the other is active (but thermally unstable; s. section 1.3. for further information about the photocycle and the differences in parent states). They absorb either at ~670 nm or ~730 nm in their Pr and Pfr states, respectively. The Pr-Pfr-transition is initiated by a covalently bound chromophore (an open-chain tetrapyrrol/bilin). Upon light-absorption, the cofactor changes its conformation, thus inducing changes in the protein environment. This ultimately activates signaling cascades through an attached catalytic module (histidine kinase (HK) in Cph1 and Agp2) which induce physiological processes. In the following, the focus will be on the two phytochromes *Cyanobacterial phytochrome* Cph1 and *Agrobacterium tumefaciens* Agp2 and both have been subject of this thesis.

The photosensory core (located at the N-terminus of the protein) and the catalytic active regulatory part (located at the C-terminus) [4],[5],[6],[7],[8] constitute the two different functional parts of all phytochromes [3]. The photosensory core consists of the PAS (PER/ARNT/SIM), GAF ("vertebrate cGMP-specific phosphodiesterases", "cyanobacterial adenylate cyclases" and "formate hydrogen lyase transcription activator, Fh1A") and PHY ("phytochrome") domains. In the case of Cph1 the HK is connected to the PHY domain. The so-called hinge region connects the photosensory core with the regulatory one in plant phytochromes. The chromophore binding domain (CBD) is the PAS-GAF region, in which the cofactor is located (see Figure 1 on the right, for an example of the CBD in Agp2). Phytochromes exist in dimeric form (homo- or heterodimerization).

X.ray crystallography helped to resolve the structure of different phytochromes [9],[10],[11]. For example the ZZZ/ssa comformation of the chromophore Phycocyanobilin (PCB) in its Pr-state in Cph1 could be confirmed [12] (s. section 1.2.). NMR analysis [13][14] similarly contributed to the valdiation of structural motives in phytochromes.

The in this work discussed crystal structures of Cph1 and Agp2 are based on the structure [10] available in the Protein Data Base [15] (code: 2VEA) and the X-Ray-studies from the group of Scheerer et al. [16] (code: 6G1Y), respectively.

1.2. Chromophore structure and binding pocket

For the following The chromophore is an open-chain tetrapyrrole (bilin, see Figure 1, left side) consisting of four pyrrole-rings, labeled from A to D (naming starts at ring A which is the protein-chromophore link). These are connected *via* three methine bridges. Different phytochromes show different prostheic groups. All of them carry the same main tetrapyrrolic spine and only differ in substitutes at rings A and D. Phycocycanobilins (PCB) and biliverdins (BV) are found in the

bacterial phytochromes Cph1 and Agp2, respectively. Plant phytochromes mostly carry a Phytochromobilin (PΦB). PCB and PΦB show an ethyl group for R_1, while BV has an additional double bond either exo- or endo-cyclic on ring A. In Agp2, the linkage of the Cys13 to BV is in the exo-form [16], although in this work an early endo-structure has been used (depicted in Figure 1, right) [16]. R_2 is in BV and PΦB a vinyl group, while in PCB it is an ethyl group.

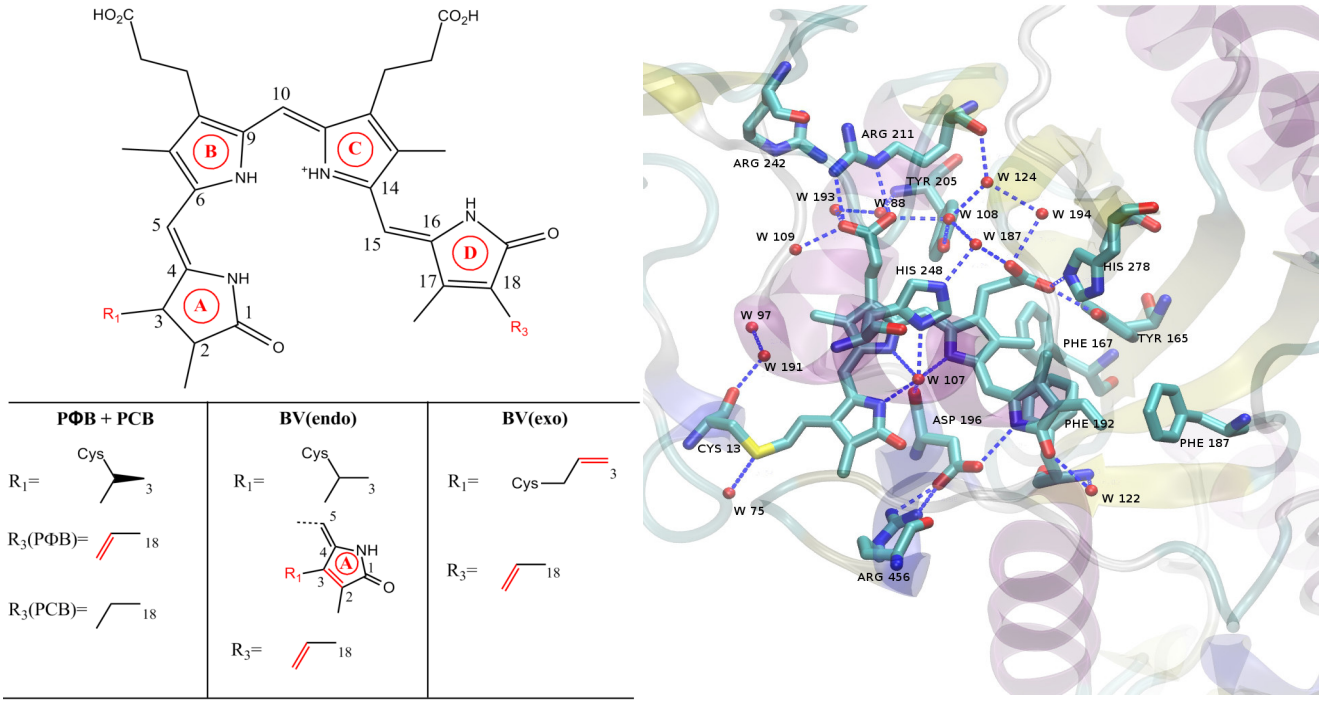

Figure 1. (left) Possible tetrapyrrolic structures for phytochrome chromophores: R_1 is in PCB and PΦB a ethyl-group with a linkage to a native Cys, in BV an additional double bond is either endo- or exo-cyclic; R_2 is in PΦB and BV a vinyl function, while in PCB it is a ethyl-group; (right) CBD of Agp2 with BV and surrounding protein residues and crystal water molecules [16]

The flexibility of the chromophore is attained to its three methine bridges which connect the four rings. Torsions around these rings represent the crucial modes of action for conversion between Pr and Pfr states. A rotation around the single-bond between rings C and D (see section 1.3. for further details) initiates conformational changes in the protein matrix, finally activating or shutting down the catalytic unit. The parent configuration of PCB in Cph1 in its Pr state is ZZZssa. Here, "Z" means "Zusammen" and "E" means "Entgegen", concerning the respective double bond configurations of the methine bridges. "s" means "syn" and "a" means "anti" and they correspond to the methine bridge single bonds. Agp2 holds the ZZEssa for the Pfr state in the crystal structure [16] (code: 6G1Y).

Furthermore, the four pyrrole-nitrogens have been found to be protonated in both Cph1 and Agp2 and in both Pr and Pfr states (Raman and NMR studies [17],[18],[19],[20],[21],[22]). It has also been found [23] that in the case of Agp2 the propionic side chain of ring C is protonated in its Pfr state. This is in contrast to observations from other phytochromes (like Cph1), where the propionates are always negatively charged and do not carry any protons [10],[13],[15].

In phytochromes, the chromophore resides inside the photosensory domain and is covalently bond to a cysteine residue. In the case of plant phytochromes and Cph1 this cysteine (Cys259) residue lies within the GAF domain, while in the case of bacterial phytochromes, such as Agp2 the cysteine

(Cys13) is located in the PAS region. In Figure 3 the CBD of Agp2 can be seen, together with its location in the photosensory core and important secondary structures (in a similar way the CBD of Cph1 is depicted in Figure 2 which is not discussed here in more detail). BV is linked to Cys13 and embedded in the protein matrix. Several residues are interacting with the cofactor. This is e.g. an aspartate lying beneath BV. It is a conserved residue in all known phytochromes and might play an important role in proton transfer processes (it is also part of the Asp-Ile-Pro motif which is conserved). An arginine, Arg211, interacting with the propionic side chain of ring B (PSC(B)) *via* formation of a strong salt bridge might also be important (see section 2.2. for further insight in the importance of PSC's of chromophores) [24],[25],[26]. Shown in Figure 1 (right) are the conserved histidine residues His248 and His278 (His260 and His290 in Cph1). Both are suggested to play important roles in the photocycle and in stabilization of the cofactor (see the results of this work in sections 9. and 10.) [23],[27],[28]. His278 directly interacts with the propionic side chain of ring C (PSC(C)) *via* a hydrogen bond. His248 interacts with the tetrapyrrole *via* a water which is bridging to the cofactor. This water is also present in all known phytochromes and lies between the four pyrrole-rings. It is hence called the pyrrole-water (p-water). This water is part of an important water-network surrounding the chromophore. The two histidines are optimal for the proton-acception and -donation events [23],[27],[28] hapening during the photocycle (s. section 1.3.). Both the named histidines and the p-water will be subject to studies in this work (parts III and IV).

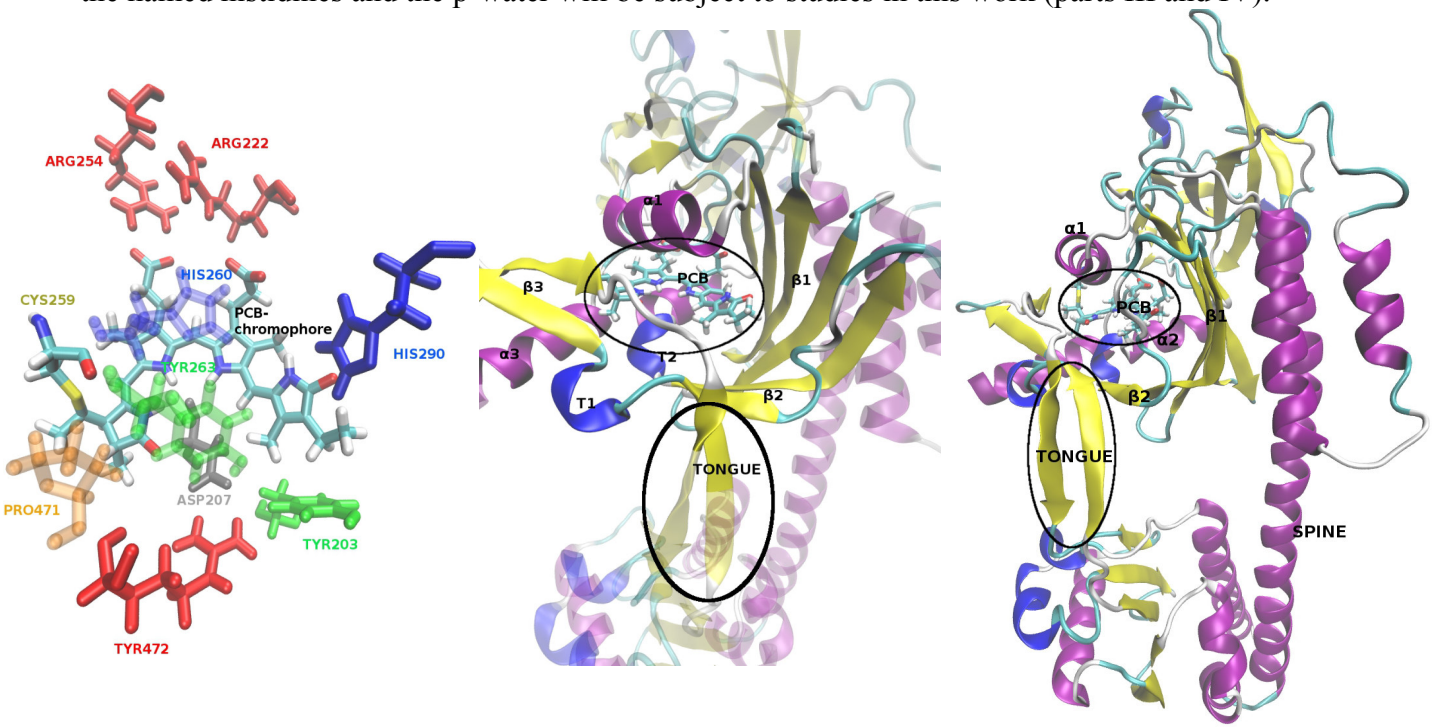

Figure 2. Crystal structure of Cph1 [10]; PDB-code: 2VEA – Pr state (left) PCB and surrounding protein residues ; (middle) CBD with important secondary structures; (right) sensoring domain

Because X-ray crystallography in general is not able to provide the positions of the hydrogens (due to low resolutions), the protonation states of these two His residues are not known. The determination of thus is not trivial and will be subject of this work (part IV). In Cph1, both NMR-studies [29],[30], as well as theoretical investigations [27], proposed a heterogeneity of the Pr-state (Pr-I and Pr-II) which differ in the protonation of His260. Newer studies by Mroginski et al. [27] suggested a cationic His260 with a N_ε-protonated His290 in Pr-I. Differing from this, in Pr-II, both histidines are protonate on N_ε. Studies by Takiden et al. [28] suggested also a possible heterogeneity for Agp1, with either both histidines either protonated at N_ε, or His250 at N_ε, while His290 is cationic. No studies on Agp2 have yet reported a heterogeneity in its parent Pfr-state.

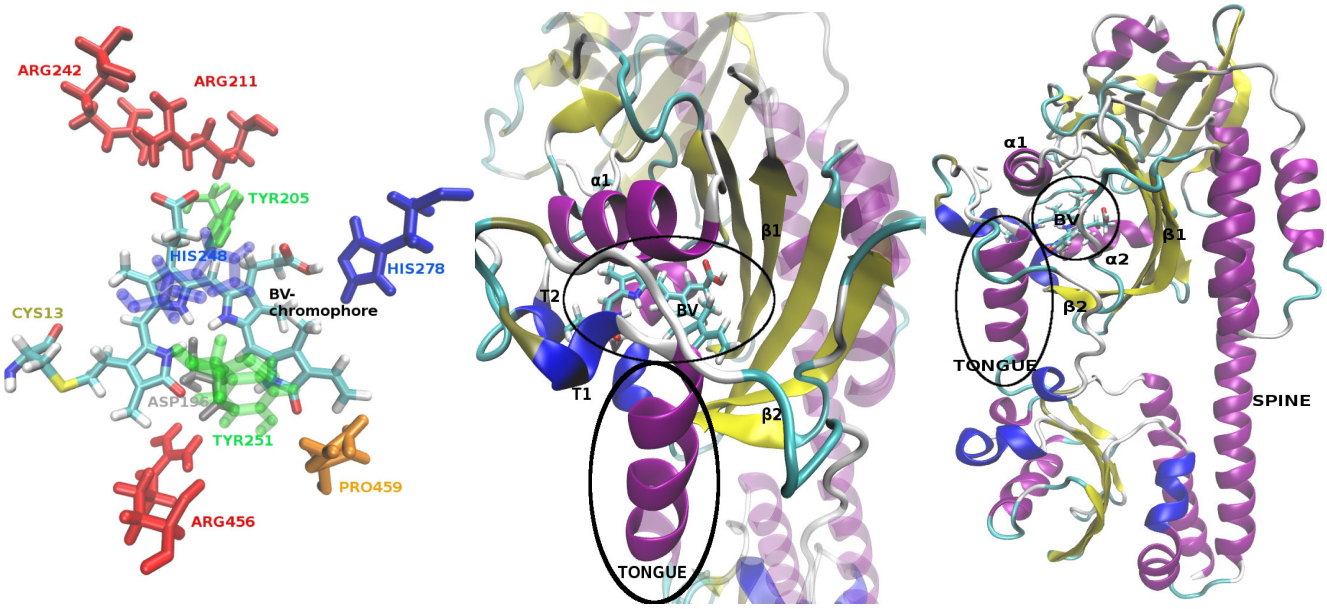

Figure 3. Crystal structure of Agp2 [16]; PDB-code: 6G1Y – Pfr state (left) BV and surrounding protein residues ; (middle) CBD with important secondary structures; (right) sensoring domain

1.3. Photocycle (comparison between prototypical and bathy phytochromes)

Photocycle. Phytochromes undergo a photocycle upon irradiation with red or far-red light. Figure 4 shows the schematic photocycle from Pr to Pfr and *vice versa*. This process is very similar in all known phytochromes [31],[32],[33],[34],[35],[36]. Red-light induces the isomerization of the C15=C16 double bond [31],[37],[38],[39] which leads to subsequent changes in the surrounding protein environment. In Figure 5 the differences between the two parent states Pr and Pfr are illustrated for Cph1 (Pr state; left) and Agp2 (Pfr state; right). ring D is twisted in BV of Agp2 in contrast to ring D of PCB in Cph1. Due to the rotation of ring D, several interactions are lost or newly gained which leads to a translation of the chromophore inside the CBD [8][40]. The photoproducts are called Lumi-R for Pr-to-Pfr- and Lumi-F for the Pfr-to-Pr-transition (s. Figure 4).

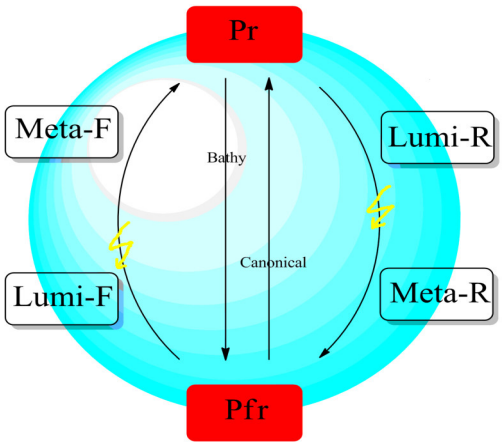

Figure 4. Schematic representation of the photocycle of phytochromes *via* Lumi-R and Meta-R or Lumi-F and Meta-F, respectively; with thermal dark-conversions for both bathy and prototypical phytochromes

The thermal relaxation of the CBD leads to Meta-R and Meta-F, respectively (Figure 4). Following this a longer Meta-F to Pr (or Meta-R to Pfr, respectively) transition takes place. Reprotonation of the chromophore, as well as substantial changes in the protein matrix occur during this step [8],[40].

Finally, the protein in its activated form is converted back to its initial Pr/Pfr state *via* far-red/red-light irradiation. Another possible back-conversion is a thermal relaxation process which is called dark-conversion, happening independent of light.

The first step in the reaction cycle is the fastest with only ~50 ps, while all other subsequent relaxation processes are in the range of milliseconds [41]. The relaxation of the CBD induces global changes of the photosensory core of the protein leads ultimately to the signaling cascades. Maybe a signaling process can only be intitiated, if both monomers in the dimeric phytochromes undergo photoisomerization processes [42].

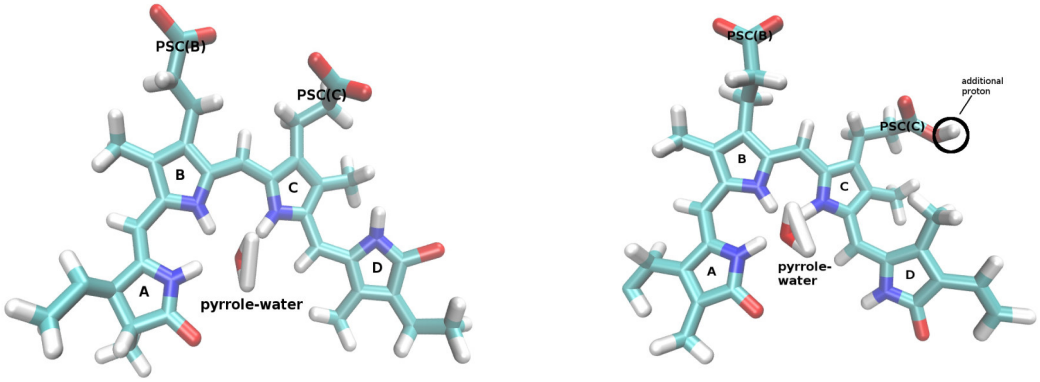

Figure 5. (left) Cph1 chromophore PCB in its Pr state; (right) Agp2 chromophore BV in its Pfr state: the ring D is tilted in the Pfr state compared to the Pr one and PSC(C) carries an additional proton; structures based on crystal structures for Cph1 [10] and Agp2 [16]

Furthermore, proton transfers, which occur during the last phase of the transition, between the Meta and Pr/Pfr states [32], might play an important role in the following signaling process. After formation of the respective Meta state, a proton (most probably from the inner pyrrole-rings) is transferred into the protein matrix. The proton-transfer might initiate further protein changes on a secondary structure level: In all phytochromes the region connecting GAF to PHY is reaching into the CBD (called “tongue”; see Figures 2 and 3, middle+right). In the Pr state it occurs as a beta-sheet (s. Figure 2), while in the Pfr state it becomes an alpha-helix (s. Figure 3) [42]. This change in structure may be induced by proton releases from the chromophore and might trigger the further movement of the phytochrome monomers, finally activating the signaling processes [42]. But this has yet to be proven. Afterwards ,the cofactor (pyrrole-nitrogen) gets reprotonated before complete conversion to Pfr/Pr. In the case of Cph1 it has been discussed that His260 might be the source for the back-protonation of the chromophore [27]. In the case of Agp2, two additional proton-involving events are assumed to play an important role: These are the deprotonation of the PSC(C) and the protonation of His278 [23] upon formation of Pr-state. All these protonation events and the rearrangements of the CBD may be the reason for the long Meta-R to Pfr transition (Meta-F to Pr, respectively) of about 300 ms.

Prototypical/Bathy phytochromes. There exist two different types of phytochromes in bacteria: The prototypical phytochromes (e.g. Cph1 and Agp1) and the bathy phytochromes (like Agp2). The main difference between both of them is the different thermodynamic stability of Pr and Pfr states. In protoypical phytochromes, like Cph1, the Pr state is inactive and the thermally stable one. Irradiation leads to conversion into active Pfr. In the case of bathy phytochromes, like Agp2, it is *vice versa*. Here, Pfr is inactive and the most stable state and far-red-light irradiation leads to the active Pr state (see Figure 5 for comparison between both chromophores in their respective parent states). In the case of Agp2, the proton at the propionic side-chain of pyrrole-ring C (PSC(C)) might

play an important role in the thermal back-reaction from Pfr to Pr. It might be transferred to His278 and this event might induce the possible dark conversion *via* a keto-enol tautomerie [23].

The photocycle is still subject to investigations. Theoretical studies might help to provide necessary information, such as protonation states and possible stuctural dynamics in the CBD to understand this process. Also prediction/computations of spectroscopic data (such as Raman-spectra) are of great importance. Interpretation and comparison to experimental resonance Raman (RR) data might provide insight into the photocycle. In this work different computational approaches are used to get insights into structure and dynamics of both Cph1 and Agp2. See parts II for theory and methods and III and IV for results for studies on Cph1 and Agp2, respectively.

Chapter 2

Rhodobacter capsulatus formate dehydrogenase(*Rc*FDH)

Protein dehydrogenases are oxireductases (enzymes) which catalyze the oxidation of various substrates, such as alcohols and aldehydes *via* the formal removal of a hydride H^--ion. These processes are often further mediated by additional cofactors, such as nicotineamideadeninenucleotide (NAD^+/NADH), flavineadeninedinucleotide (FAD+/$FADH_2$) and further cofactors (see section 2.2.). One example is the human aldehyde dehydrogenase [ALDH2] [43] which is important for the degradation of alcohol. It mediates the reaction RCOH + NAD^+ + H_2O → RCOOH + NADH + $2H^+$ [44]. Without a correctly functioning aldehyde dehydrogenase, the organism is not able to convert the toxic aldehydes into non-toxic acetates $RCOO^-$. Such is also called the Oriental-flushing syndrome [45]. The ALDH2 contains additional to NAD^+ (the electron-acceptor) a Mg-containing cofactor mediating the oxidation of the aldehyde into the carboxylic acid. Furthermore, different protein residues play an important role in stabilization and catalytic conversion of the substrate (such as Cys [45]).

Formate dehydrogenases (FDHs) are a member of the dimethyl sulfoxid (DMSO) reductase family. They mediate the oxidation of formate ($HCOO^-$) to carbon dioxide CO_2 *via* cleavage of a C-H bond, The electrons are donated to NAD^+ or cytochrome (CYT) [46], depending on the organism: Aerobic metabolisms, e.g. yeast and bacteria, contain NAD^+-dependent FDHs, while anaerobic prokaryotes contain CYT. It is a key process for the organism (e.g. bacteria like *Rhodobacter capsulatus*) to obtain energy and reducing-equivalents. Metal-dependent FDHs contain a metal-containing cofactor with either molybdenum (Mo) or tungsten (W) [47] at their active-sites. Furthermore, they all show conserved Cys (native form or seleno-Cys), His and Arg residues in near vicinity of the molybdenum-containing cofactor (Moco). In the following, more detailed insight into FDH of *Rhodobacter capsulatus* (*Rc*FDH) is given, since this work is focused only on this FDH. FDHs of other species, such as from *E. coli* are only mentioned but not discussed in detail. In the following the structure and function of the Moco and other cofactors, such as iron-sulfur clusters are described in greater detail. Many of the following details (such as Moco-structure) are conserved between known FDHs.

2.1. General protein structure motives and global mode of action

*Rc*FDH is a Mo- and NAD^+-dependent oxygen-tolerant enzyme. Its functional unit consists of a dimeric heterotrimer [48] (shown schematically in Figure 6 on the left). The $(\alpha\beta\gamma)^2$-unit holds the active site α (FdsA) with the Moco and five different iron-sulfur clusters (FeS), namely four Fe_4S_4 clusters and one Fe_2S_2 cluster. The γ-unit (FdsG) contains only a Fe_2S_2 cluster. And finally the β-unit (FdsB) contains one Fe_4S_4 cluster, as well as one flavinemononucleotide (FMN) cofactor and the NAD^+ binding-site. The formate oxidation ($HCOO^- \rightarrow CO_2 + 2e^- + H^+$) occurs at the active site located in the α-monomer and at the Moco-site. The electrons are then transmitted *via* reduction of the metal (here Mo): Mo(+VI) → Mo(+IV); into the Moco. The protein and Moco exist therefore in at least two different redox states (more on possible oxidation states and the possibility of a MoV-state in sections 12. and 13.4.): one with Mo(+VI) and one with Mo(+IV). Afterwards, the electrons might be translocated into the protein matrix *via* reduction of the different FeS clusters, until they finally reach the FMN-site. Here FMN catalyzes the reduction of $NAD^+ + 2e^- + H^+ \rightarrow$ NADH. It has also been found that *Rc*FDH is able to catalyze the reverse reaction, namely the CO_2-reduction to formate [48]. The observed reaction rate for formate oxidation was $k_{cat} = 2189\ min^{-1}$, while the one

of the backreaction was k_{cat} = 89 min^{-1} [48]. The CO_2-reduction is very slow, because the electrical potential of the FeS clusters might hinder the electronic flow from flavine to the Moco-site [48].

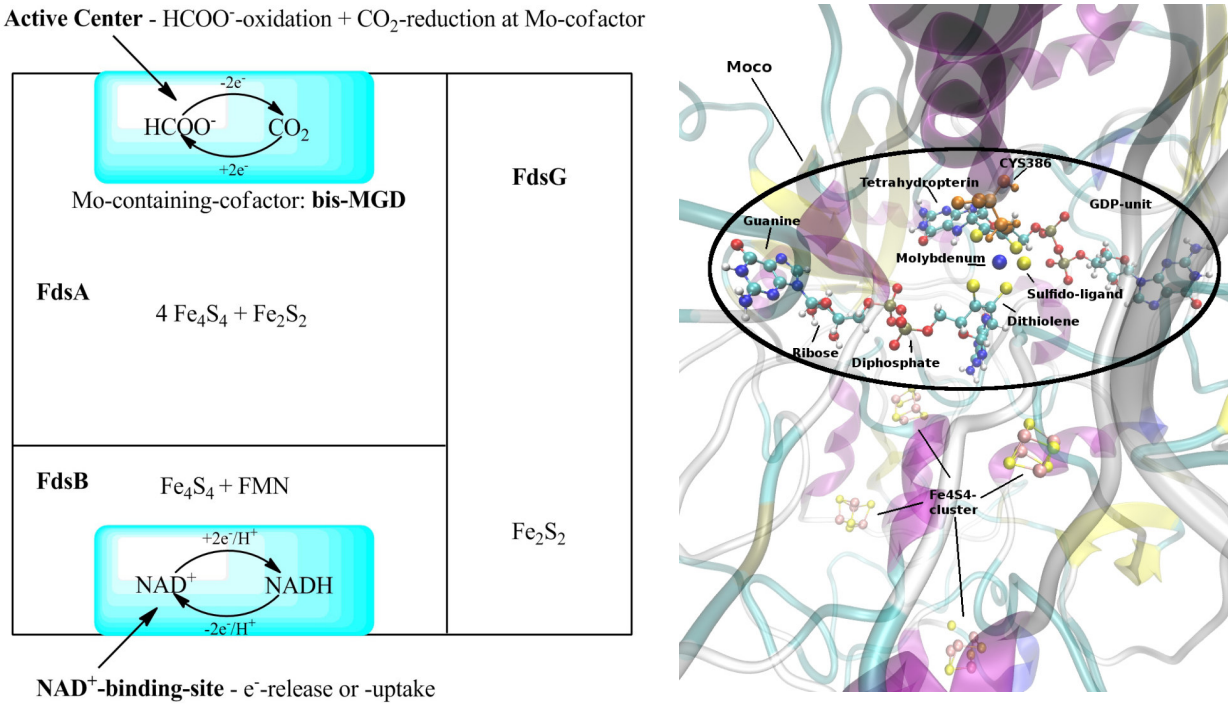

Figure 6. (left) Schematic representation of the *Rc*FDH heterotrimer and the corresponding cofactors and its active sites (Fig. inspired by Fig. 1C in [48]); (right) *Rc*FDH homology model [49]

Until now, five different genes are known to code for *Rc*FDH. These are FdsA, FdsB and FdsG (responsible to the α-, β- and γ-units). Additionally, there exist two other genes FdsC and FdsD which have been recently investigated [48]: The former is an analog to the FdhD gene in FDH of *E. coli* and is important for sulfuration of the Moco (see section 2.2.) [48], while the latter has no known analog in *E. coli* and is unique to *Rc*FDH. It seems to be important for insertion and stabilization of the Moco inside the protein matrix, while bound to FdsC [48]. It also might be restricted to oxygen-tolerant FDHs and might be important for the protection of the Moco from oxygen-induced degradation processes. Thus, both are necessary for maturation processes of FDH and cofactor-insertion [48].

Until now, no crystal structure is available for *Rc*FDH. The above mentioned results were due to spectroscopic [50] and kinetic studies [48][49]. For the here described theoretical studies, a homology model [49] has been taken as a starting point. This has been built from the crystal structures of *E. coli* (see section 5.2.1.).

2.2. Molybdenum-containing cofactor

Several cofactors are incorporated into the protein, such as FeS clusters, FMN and also the Moco. Since this work focus' only on the Moco and surrounding protein residues, the other cofactors will not be discussed further.

The Mo-containing cofactor, the Moco, consists of a metal (here molybdenum) ligated by two dithiolene moieties. These are each covalently attached to a 5, 6, 7, 8-tetrahydropterin ligand on one side and a guanosine diphosphate (GDP) on the other side. Hence, the Moco is called bis-molybdopterin guanine dinucleotide (bis-MGD). In Figure 7 one possible Moco-structure is shown.

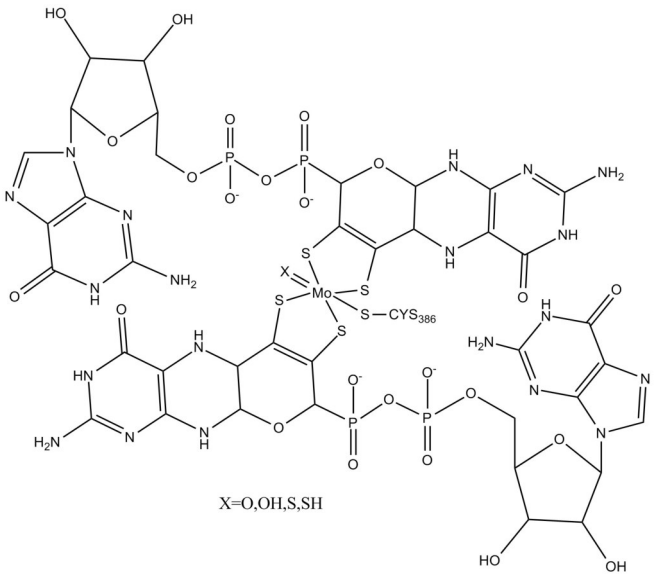

Figure 7. Structure of the oxidized bis-MGD cofactor (Moco). X means potentially binding of S, SH, O or OH ligands and Cys-ligation might be optional, as discussed in the text.

The dithiolene moieties may be important for tuning the redoxpotential of Mo [51]: They are able to adjust to the oxidation state of Mo (as they have different oxidation states themselves, shown in Figure 8). Thus, they might participate in the redox process at the Mo-site. Furthermore, by bending and thus changing their geometry, they can also possibly influence the electronic environment of molybdenum. Stronger folding of dithiolenes with respect to the metal seemingly facilitates higher electron donation potential in Mo [51]. Therefore, for high oxidation-states of molybdenum such strong distortions have been found [51].

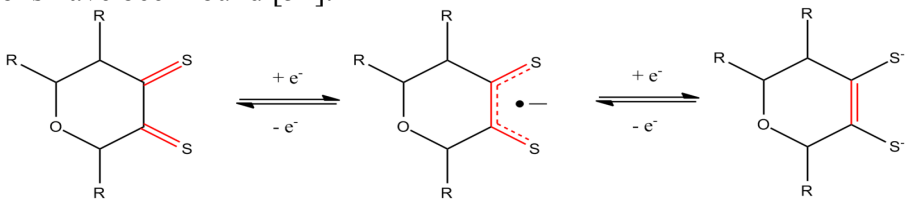

Figure 8. Dithiolene ligand and its different redox states (picture was inspired by Fig. 2 in [51])

The pterin ligands may also be important for modulating the redoxpotential of Mo [51] or for the e^--transfer from the active site to the interior of the protein. This is due to the fact that they also exist in different oxidation forms, as shown in Figure 9. The electron-deficient pterin ligand might facilitate electron-withdrawal from the dithiolene ligands and ultimately from molybdenum [51]. Furthermore, strong hydrogen-bonding interactions between pterin groups and the protein environment have been observed for molybdo-enzymes [51]. These might further play an important part for redox processes, coupling electron-transfer-pathways with proton-translocations [51]. Finally, the protein-pterin interactions might be important for stabilization of the energetically less efficient distorted trigonal prismatic sturcture of the Moco inside the binding-site [51].

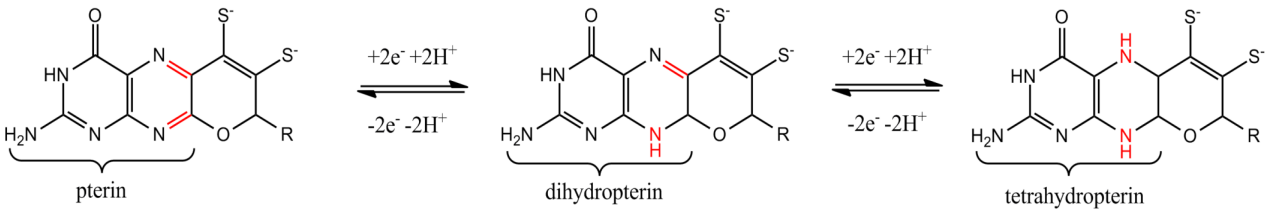

Figure 9. The opterin ligand and its different redox states (picture was inspired by Fig. 4 in [51])

In addition, the pyran ring linking the pterin group to the dithiolene ligands might exist in an ring-open and closed form and might also be important for redox reactions [51].

MoVI has two additional possible ligation-sites: In *E. coli* it has been found that one is occupied by a seleno-cysteine (Se-Cys), while the last ligand is a sulfido or thiol-group [52]. Since no crystal structure is available for *Rc*FDH only assumptions (based on spectroscopic and kinetic studies of *Rc*FDH and of other FDHs, since all carry similar motives) can be made of what additional ligands may be present in both MoVI and MoIV-forms (see Figure 10). It has been suggested [48],[49],[50], [52],[53],[54],[55] that in the active form of FDH, the oxidized molybdenum (MoVI) is ligated to an additional native (Se-)Cys386 and a S*-ligand (see Figure 10, left and middle). Therefore, the oxidized Moco with MoVI might be hexacoordinated, while the reduced form may be pentacoordinated [49],[50],[52],[53],[55] and only carrying a S*-ligand and no Cys (see Figure 10, right). This is controversial, since Maia et al. [54] also suggested based on EPR-results that the native Cys is not leaving the ligation site upon Mo-reduction and instead stays coordinated to the metal in its MoIV-form. They also suggested the presence of a (at least temporary) SH-ligation at the metal. Therefore, it remains unclear, whether Cys is ligated to the MoIV and what other ligand may be present. For more details see section 3.3.). Another question is, whether or not the substrate (formate) is directly binding to molybdenum which will be possible, if either the Cys or the S*-ligand is leaving the coordination site [48],[49],[50],[52],[53],[54],[55],[56].

Furthermore, in its oxidized form also an inactive species of *Rc*FDH could be observed [50]. Spectral analysis [50] suggested that this form carries either an O-ligand or and OH-ligand at the MoVI-site, instead of the sulfido-group and represents also a hexacoordinated MoVI-species. Therefore, it is called the desulfo form (or prestate). This species has been observed, upon suppression of the FdsC gene [50].

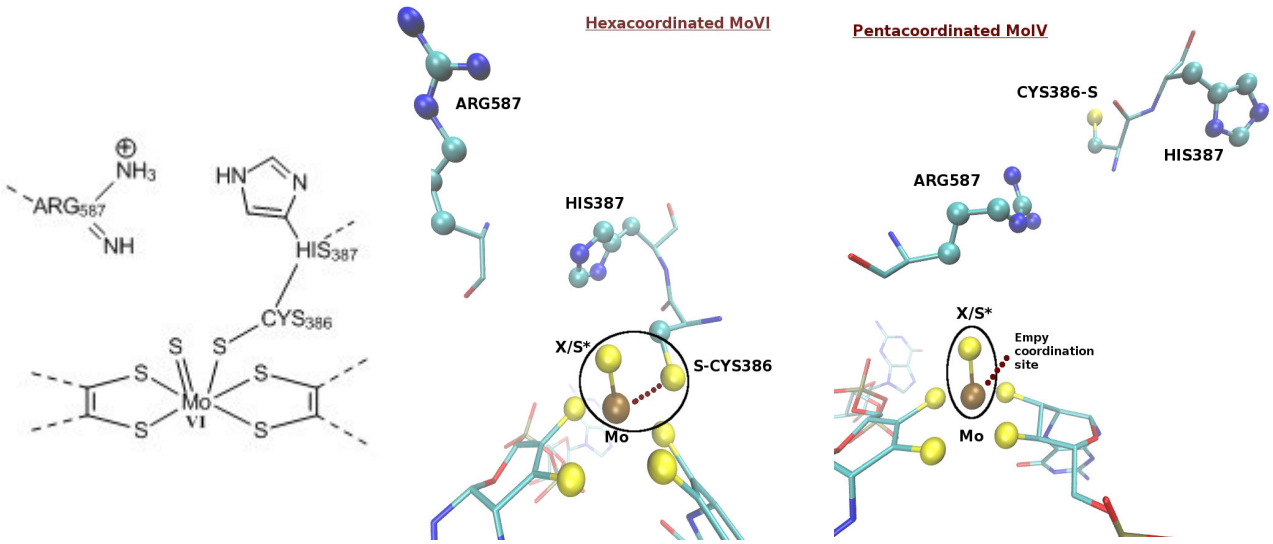

Figure 10. The Moco in *Rc*FDH and nearby conserved protein residues Cys, His and Arg in FDHs in possible ligation scenarios of molybdenum (regarding S*- and Cys-coordination of Mo): (left) schematic representation of all-sulfur hexacoordinated MoVI; (middle) oxidized *Rc*FDH and (right) reduced *Rc*FDH; in the latter two, the structures of the homology models of *Rc*FDH by Hartmann et al. [49] are shown; also the conserved residues Arg587 and His387 are displayed: X denotes either S, SH, O or OH-ligand (pictures inspired by Fig.1 in [53])

As shown in Figure 10, three conserved protein residues, namely Cys386, His387 and Arg587 can be found in direct vicinity of the Moco. These have been attributed different functions [48]:

According to Hartmann et al. [49] Cys is important for the proton abstraction from formate and thus directly involved in the catalytic process. Therefore, it has to be detached and deprotonated (normally in native environments Cys is protonated due to its high pKa) and both Arg and His play important roles in this. His tunes the pKa-value of Cys, while Arg stabilizes its charge *via* salt-bridge formation. Arg is also important for correct insertion and subsequent stabilization of the substrate inside the binding pocket at the Moco. This has been found in mutation studies [49].

2.3. Possible mechanistic properties at the active site

As can be seen, the structure of the Moco is still area of discussion. For further information of possible involvement of the Moco in the reaction cycle and potential mechanisms see section 3.3. There, the underlying motivation of the study of Moco-structure and its mechanistic properties, as carried out in this work, are discussed in further detail.

Chapter 3

Motivation for computational investigations of large biomolecules

Phytochromes, as well as FDHs have different areas of interest. For example, it is still not completely clear how phytochromes interconvert between their two parent states on an atomic level. Intermediate steps of the photocycle have to be clarified, as well as how changes of the chromophore finally induce changes on a macroscopic scale. It is important to investigate the role each residue plays in the photocycle, as well as possible interactions with the chromophore. For this, computational studies combined with experiments are of importance. This includes the development of new and improved methods for simulating the native environments. The same goes for the study of FDH. Structure of the molybdenum-containing cofactor, as well as mechanistic properties of it and the involvement of surrounding residues His, Arg and Cys are still area of discussion. They may be estimated with the help of computational methods, such as density functional theory (DFT). In the following chapter the motivations and concrete aims for the individual projects of this thesis are shown and discussed.

3.1. Phytochromes – Improvement of available methods

Classical molecular dynamic (MD) simulations or force field methods (for theory see section 4.1.) provide a tool for getting access to conformational states and dynamics of macromolecules. It has many advantages, like high simulation times at relative low computational costs which makes it a widely used method to examine biological matter. On the minus side, possible problems can arise from shortcomings of classic force field (ff) methods: In past studies [57],[58],[59],[60],[61],[62], [63],[64] it has been indicated that the point-charges as included into classic molecular mechanic (MM) MD-simulations are not able to describe the water-network inside biological probes realistically. This was especially true for explicit water molecules around cofactors inside e.g. proteins. A possible reason for this may be the insufficient description of the charge-environment with point-charges. In case of Cph1, the charges for individual protein atoms included in the used force field are obtained from experiments and calculations of the apo-protein. Hence, in the absence of the cofactor. Electrostatic effects of the chromophore on the surrounding protein residues and *vice versa* are therefore unaccounted for.

MD-simulations of Bacteriorhodopsin (BR) as carried out in the past by the group of Eichinger et al. [57] and others [58],[59] showed a decreased stability of the structure of the water network surrounding the cofactor. Furthermore, important hydrogen bonds, such as between a crystal water and the Schiff base are disrupted. This led to the formation of a hydrogen-bond between cofactor and asparagine85 in MD-studies of Grudinin et al. [62]. Additionally, the complete water network showed more fluctuations, as observed by Kandt et al. [61]. In contrast to this, NMR-findings [60] predicted a stabilized water- and hydrogen-bonding-network around the cofactor.

Kandt [61] and Grudinin [62], as well as Hayashi et al. [65] attributed the unstable water-network and the high fluctuations of individual water molecules to the lack of polarization effects in the applied force field. The simple point-charges were not able to reproduce the native structure and hydrogen-bonding network, especially concerning the Schiff base.

In addition to the development of polarizable potential functions by Rick et al. [66] and its application by Hayashi et al. [65], Babitzki et al. [67] addressed the instability of water molecules in BR *via* the application of a so-called polarized force field. Hybrid quantum mechanical/molecular mechanical (QM/MM) methods have been used for the development of the latter. MD-simulations of BR verified the stabilization of hydrogen-bonding and water-networks, when using the new polarized force field [67].

In studies of Mroginski et al. [63] and unpublished studies by Daminelli et al. [64] similar observations have been made concerning MD-simulations of the sensory domain of Cph1. They showed a decay of the structural stability inside the chromophore binding pocket: The pyrrole-water did not stay stable and left its crystallographic position during the simulation. Hydrogen-bonding interactions between the crystal water and the chromophore, as well as surrounding amino acids are disrupted or lost. This might be, because the applied point-charges for the chromophore and surrounding protein residues did not sufficiently describe the real electronic interactions. These findings are in contrast to crystallographic studies [10].

Furthermore, snapshots of these trajectories were used for Raman-spectra computations [63]: The water-fluctuations in the simulations led to a broadening of the calculated Raman-spectra bands [63]. Especially the bands connected to the dynamics of the pyrrole-water were affected. The NH-in-plane (ip) rocking of the pyrrole-nitrogen at ring C of the chromophore is coupled to the pyrrole-water. The high mobility of the latter lead to a broadening of the corresponding band in the computed Raman-spectra [63], while experiments show a sharp peak at 1570 cm^{-1}. This makes the computed spectra less comparable to experimental ones and thus less helpful in interpretation of the latter.

Since the pyrrole-water might play an important role in the photocycle, its correct treatment in MD-simulations, as well as Raman-spectra calculations is of high importance. A more realistic treatment of the charges would improve the comparability of theoretical data with experimental studies, such as resonance Raman (RR) spectroscopy. Therefore, the instability of pyrrole-water has to be addressed. A plausible solution is the incorporation of polarization effects of the cofactor on the surrounding protein residues and *vice versa via* a polarized force field.

In this work several MM simulations of the photosensory domain of Cph1 have been carried out with both classical MM force fields and also with a newly generated polarized force field for the Cph1 PCB-binding pocket. Additionally, Raman-spectra have been calculated for each type of simulation and then compared to experimental studies. This was done in order to investigate and address the band-widening problem of the NH ip-rocking of ring C. This study will demonstrate a better treatment of the charges in near vicinity of the chromophore, thus generating more realistic and stable conditions during MD-simulations of Cph1. Additionally, it will be displayed that such a force field generates more trustworthy visualizations of the dynamics inside the CBD and more realistic computed Raman-spectra. This will be helpful in interpretation of experimental RR studies of phytochromes. It will provide a helpful tool in future investigations of the photocycle with a combined theoretical/experimental spectroscopic approach. The theory and procedure for the generation of the polarized force field, as well as details of the application are shown in sections 4.1.5. and 5.1.2. The results of these studies are shown in part III.

3.2. Phytochromes – Determination of structural and dynamical information.

Determination of the crystal structure of biomolecules is always an important part of understanding the structure and function. But even with those results, the understanding of the molecular structure

still has to be improved, since the resolution of X-ray crystallography is often to low to visualize the hydrogen (H) atoms bound to the protein structure and possible cofactors. H atoms might play an important role in protonation events, such as has been suggested in the photocycle (see section 1.3. for further details) of phytochromes [23],[27],[32],[42]. Therefore, it is necessary to use theoretical methods to determine the protonation states of the various amino acids and the chromophore. This becomes even more important in the case of titratable amino acids. These can either be protonated or not, influenced by their immediate environment and possible interaction partners. Under physiological ph-conditions most amino acids, which are not in close contact to cofactors, are treated as charged (H-acceptor groups are protonated, while donors are deprotonated). It becomes problematic for amino acids in near vicinity of a cofactor (such as chromophores like BV), since it influences the pKa values of the surrounding titratable groups. The most difficult is the determination of the protonation states of histidine residues (for they have two possible H-acceptor sites, i.e. nitrogens (N), leading to three different protonation patterns: H on N_δ ,named d-protonation; H on N_ε ,named e-protonation; or H's on both N's, named p-protonation; see Figure 11). Here, the correct protonation state has to be evaluated out of these three possibilities. In all phytochromes two histidine residues in near vicinity to the chromophore are conserved (in Agp2 they are named His248 and His278). It is assumed that they play a crucial role in stabilization of the chromophore in the CBD and , additionally, may affect the conversion from Pfr to Pr and *vice versa via* acceptance and donation of protons during the conversion [27]. In Agp2 these residues are located directly above the BV (His248) and in the case of His278 near the propionic side chain of ring C (PSC(C)). To investigate their role in the stabilization of the CBD it is important to find the correct protonation state of both residues beforehand. It has recently been shown in the case of Cph1 [27] and Agp1 [28] that different positioning of hydrogen atoms on the corresponding His might induce structural changes and instabilities on the chromophore and its protein surroundings.

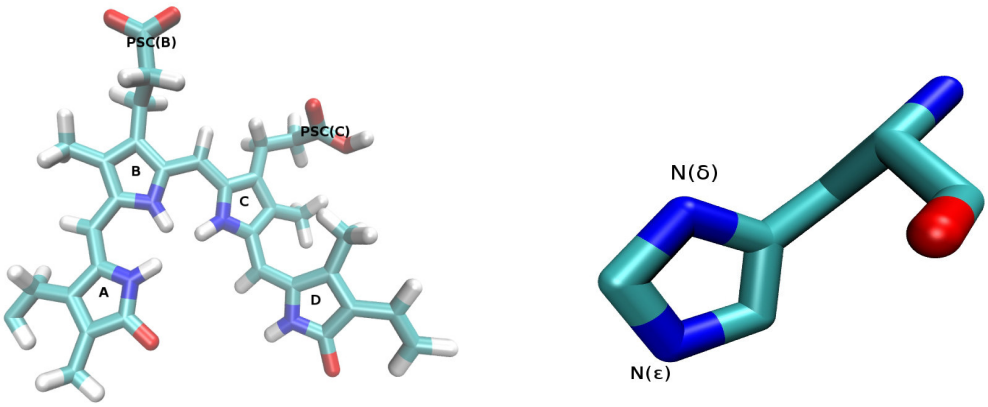

Figure 11. Structure of Agp2 chromophore BV with protonated propionic chain of ring C (left) + possible protonation sites of histidine side chain (right); based on crystal structure of Agp2 [16]

The computational investigation issuing the protonation state of His248 and His278 in Agp2 is carried out *via* QM/MM computations (see section 5.1.3. for more insights into the applied methods) of nine different models. These are taking all possible combinations of protonations of the two N's of both His into account: Namely, these are dd, de, dp, ed, ee, ep, pd, pe and pp (first letter refers to protonation-state of His248, second letter to His278). Subsequently, the computed models have been compared with the X-Ray structure [16]. The recent discovered X-Ray structure [16], which is the first for Agp2 in its Pfr state (beforehand the only available crystal structure of a bathy phytochrome in its Pfr state, was that of Pseudomonas Aeruginosa [11],[68],[69],[70]), is used as reference structure in this study. The comparison of stabilities of the different computed models

compared to the X-ray structure was the most critical testing parameter. It has been used together with the stability of the hydrogen-bonding-network (H-bond-network) inside the CBD and potential hydrogen transfers from the histidines to the cofactor in order to the determine the most probable protonation state of His248 and His278. The results of these studies are shown in part IV, section 9.

These results provide a good starting-point for further investigations of the CBD of Agp2. The propionic side-chains of BV may play an important role in the photocycle and stabilization of the cofactor in its pocket [24],[25]. This has been already investigated for BV-containing phytochrome DrBphp and PCB-containing Cph1. Lagarias et al. [24],[25] studied the influence of amidation of PSC's in the chromophores of named phytochromes on the formation of stable parent states and the photocycle itself. They found [24],[25] that in DrBphp, the modification of PSC of ring B (B) led to differences in the Pr state (compared to the native Pr state) and multiple Pfr states. PSC(C) amidation led to dual-Pr states and subsequently to dual-Pfr states. They concluded that PSC(B) may be responsible to pKa-tuning of the cofactor and its surrounding residues. Further it seems to play an important role in red-light sensing. PSC(C) defines the width of the detected spectral region. Their studies of Cph1 led to similar conclusions concerning PSC(B), whereas modification of PSC(C) inhibited completely the formation of a stable Pfr-state.

These findings suggest that PSC(B) and PSC(C) play an important role in both structure and dynamics of phytochromes. With the help of the correct protonation state of His248 and His278, the influence of both PSC(B) and PSC(C) on the dynamics inside the CBD of BV has been investigated for Agp2. Different computational models, in which the PSC's have been methylated (either only one of them or both) were computed. The impact of these structural changes on the stability of the CBD (compared to dynamics of the native protein and to the crystal structure of Agp2) has been evaluated and used to interpret the corresponding RR experiments [26]. The latter findings of Hildebrandt et al. [26] suggest that one monomethylated variant of BV remains photoactive, while the other leads to an abolished photocycle. These investigations will lead to a better understanding of the roles of PSC's of BV in Agp2 and in phytochromes in general. Furthermore, the obtained results of these studies will provide helpful tools for future investigation of the photocycle of phytochromes. The results are shown and discussed in section 10.

3.3. *Rc*FDH – Structural and mechanistic questions

FDHs are of high interest for both industry and environment. Since they are able to catalyze both the oxidation of formate and the reduction of carbon dioxide, they provide the necessary tools for generation of C-building blocks out of gaseous CO_2. Formate might function as a reduced carbon (C)-source for further modifications and can also be used as a solid energy source. Simultaneously, the amount of CO_2 in the atmosphere could be reduced, helping to reduce the Greenhouse effect and global warming. The key for application of FDHs and artificial enzymes with similar function for these purposes, is the complete understanding of the reaction cycle of FDHs and its catalytic activity. Only then is it possible to create and tune the desired CO_2-reduction activity. Prerequisite for this, is the correct determination of the structure of involved cofactors.

Several studies have been carried out, addressing the correct ligation of the molybdenum-site of the Moco in FDHs. They commonly agree in four ligation-sites being occupied by two dithiolene moieties (see Figure 10). But further assumptions differ from work to work. In X-Ray-studies on FDH of *E. coli* (*Ec*FDH), a Selenium(Se)-Cys has been found to be ligated to the molybdenum [52], while a sulfido- or thiol-ligand is the sixth ligand in the oxidized form of the Moco. Mota et al. [54] suggested also a six-coordinated Mo in its oxidized form (with Se-Cys- and S*-ligands), but only as an inactive prestate which leads to a pentacoordinated active form, lacking the Se-Cys coordination.

Schrapers et al. [50] found in spectroscopic studies of *Rc*FDH two different oxidized species which both showed hexacoordination (see Figure 10, left), with native Cys386-ligation and either S*-ligand (active) or possible O- or OH-ligation for the inactive species. Leimkühler et al. [48],[49] followed this approach for *Rc*FDH. It is not completely clear what ligand is present in addition to the two dithiolenes in the reduced form of FDH. A sulfido-ligand as fifth ligand has been mostly assumed (see Figure 10, right), but never completely verified. It has been assumed that maybe the (Se-)Cys-ligand is dissociating from Mo [49],[50],[52],[53],[55],[71]. The thus generated free coordination-site might be occupied by the substrate formate [49],[50],[52],[53],[55],[71]. This is currently still under discussion. For example Maia et al. proposed a hexacoordinated MoIV with a Cys-ligation in addition to the formation of a SH-group ligated to the molybdenum [54].

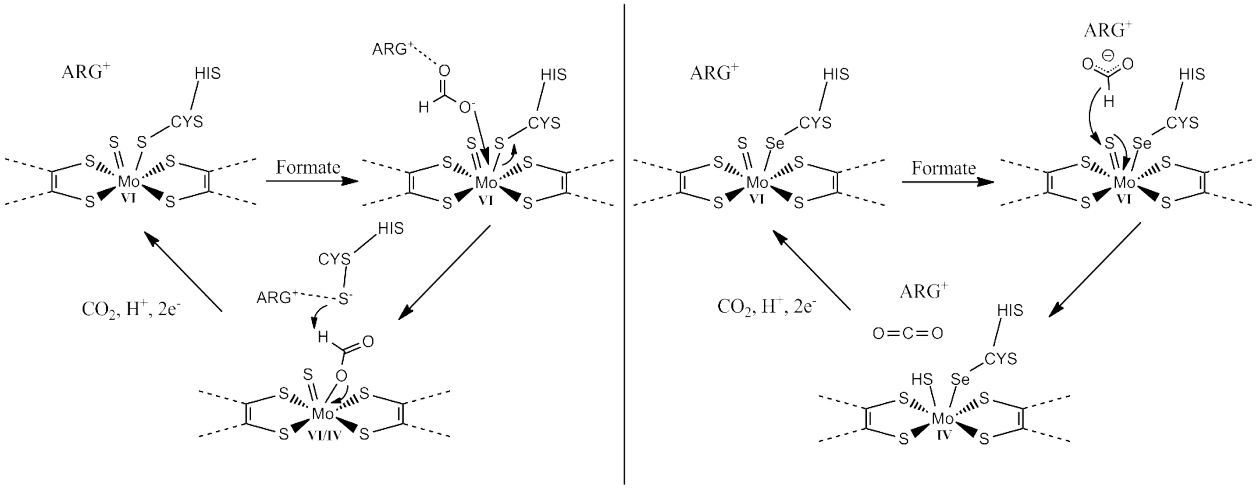

Figure 12. Different catalytic reaction mechanisms (left) proposal of Hartmann et al. [49] with direct formate binding and H-abstraction by a dissociated Cys-ligand; (right) proposal of Maia et al. [54] with formate residing unbound in the second coordination sphere and proposed hydride transfer from formate to sulfido ligand (the pictures were inspired by Fig. 7 in [49] and Fig. 6 in [54])

The disagreement of the correct ligation-sphere in molybdenum-containing FDHs has a huge impact on the understanding of the catalytic cycle, since it depends strongly on the correct positioning of the ligands. Until now there have been several mechanistic proposals, some of which are presented here:

Concerning a hexacoordinated MoVI as starting point in the catalytic cycle, several studies assume that the Cys (ligated to MoVI) leaves the metal center upon formate approach, while the sulfido-/thiol-group stays at the metal [50],[52],[71]. Formate then directly ligates to the molybdenum-center. Here, the Cys386 moiety is mediating the H-abstraction from formate and thus the conversion of formate to carbon dioxide. Another study proposes the dissociation of the sulfido-atom instead of the Cys-ligand from the MoVI site [56]. In both the leaving group returns to the metal after the cycle is complete. A second common characteristic of both is that the sulfido-/thiol-group is not involved in the oxidation of formate or subsequent electron translocations. These mechanisms are based on the crystal structure of the FDH of *E. coli*. The second reaction proposal [56] has assumed to be false after revision and reinterpretation of the X-ray data [52].

A third possible reaction cycle states that a hexacoordinated MoVI center is inactive and has to be activated in a first step [53]. Here, the Cys386 is leaving the ligation site *via* temporary formation of a disulfide bridge with the S*-ligand. This generates a free coordination position for formate which is directly binding to the Mo [53]. This is hence called the sulfur-shift mechanism [53].

Two approaches (for either *Rc*FDH [49] and *Dd*Fdh [54], respectively) are currently discussed in the scientific community and are of major interest: Namely the ones from Hartmann et al. [49] and from Maia et al. [54] (which is a refined version of the already known mechanism from Niks et al. [72]). Both are shown as a schematic in Figure 12.

In the mechanism proposal from Hartmann et al. [49] the process starts with a hexacoordinated MoVI-species (2 dithiolene, 1 Cys386 and 1 sulfido ligation). Upon the arrival of formate, Cys is leaving the ligation site, while formate directly binds to the metal. C-H bond cleavage is done by Cys abstracting the H^+ from formate, while two electrons are transferred from the substrate to MoVI, thus changing its oxidation state to MoIV. In a final step CO_2 is released, the proton further translocated into the protein matrix and the two electrons directed to the FMN-site (most probably *via* the FeS clusters), oxidizing MoIV back to MoVI. Simultaneously, Cys is returning to Mo and the cycle can restart. Site-directed mutageneses studies [49] revealed the important role of the His387 and Arg587 residues which are in near vicinity of the cofactor (similar residues are conserved in all known FDHs, thus underlining the importance of these for catalytic activity): While Cys386 is important for proton uptaking, it has to stay deprotonated which under physiological pH is nearly impossible (due to its high pKa). A higher ph-optimimum (ph=9) for the catalytic reaction in *Rc*FDH has been found [49], ensuring a deprotonated Cys. His387 further was proposed to mediate the pKa of Cys386. Arg587 is directing and stabilizing the substrate to its correct position at the metal center. Further it helps Cys386 to stay charged *via* formation of a salt-bridge with it. In the tested variants, in which the named residues have been substituted to a residue with differing functionality at the side-chain (e.g. unpolar group), the enzyme looses its catalytic activity [49].

The proposal from Maia et al. [54] goes a different way: They also started with a similarly hexacoordinated Mo species. In their studies on *Dd*FDH, they found that in the presence of cyanide the enzyme becomes inactive. Their interpretation is that CN^- forms a thiocyanide group with the sulfido-ligand. And since this renders the whole oxidation process obsolete, they concluded that the S*-ligand has to be important in the cycle and has to be directly involved. They further suggested that on one hand, Cys is not leaving its coordination position and plays no direct role in proton release and uptake. Thus, formate is not directly ligated to the metal center, only approaching and stabilized by an Arg residue. A hydride ion is then abstracted by the sulfido-group (MoVI becomes MoIV). This leads to the release of CO_2 and the formation of a SH-ligand. After transferring the electrons deeper into the protein, a proton is released and SH becomes a S-ligand, while MoIV becomes MoVI again and the cycle can start from the beginning. These findings they supported by kinetic studies (which also suggest an activation process of unknown nature [54]) and by EPR measurements [54] which show the presence of a coupling H^+ in the second ligation sphere of Mo. Further they state that no experimental evidence, besides XAS data [50] is given that Cys is actually leaving the reaction site. Nevertheless, they do not exclude the possibility of a pentacoordinated MoIV state and the involvement of a dissociated Cys in the translocation of a proton [54].

A third recent publication by Robinson et al. [55] shows a combination of certain aspects of these two mechanisms: Starting with a hexacoordinated MoVI in *Ec*FDH they proposed a direct coordination of formate to the molybdenum-site after dissociation of cysteine, based on inhibitor-studies and electrochemical experiments. Furthermore, they suggested that the sulfido-ligand abstracts either a proton or a hydride from bound formate, thus reducing MoVI to MoIV and releasing CO_2. Temporarily a SH-ligand is formed which returns back to the sulfido-form after releasing a proton into the protein matrix and restart of the cycle.

All these mechanistic suggestions have reasonable ideas, underlined by experimental evidences. Unfortunately they differ often in important key aspects, such as the release of Cys and the direct

ligation of formate. Also the proton abstraction from formate and the electron transfer from formate to molybdenum is not consistent in all named publications. The question remains which one of these proposed reaction cycles is correct or whether it is a combination of different elements of several of these mechanisms. Furthermore, these proposals lack some exclusive things: For example, the mechanism from Hartmann et al. [49] does not include the EPR findings, suggesting the hydrogen in the second coordination sphere. In the Maia-mechanism [54], a controversially issue is the inhibition by CN^-. They attributed it to the formation of SCN^- with the S*-ligand. This might be wrong. CN^- is known to be a very strong ligand with a strong ligand field splitting (creating low spin complexes by formation of π-backbinding) and it is possible that CN^- simply initiates the leaving of all ligands from Mo and exchanges them, because it forms the more stable complex $[MoVI(CN)_6]^0$ which would also inactivate the enzyme. Another point is the Cys-dissociation from the metal observed in known crystal structures of different FDHs in the reduced state [52],[53]. There, Cys is dissociated from MoIV. Additionally, the EPR-findings of an intermediate MoV-species (which would suggest a sequential electron release/uptake in contrast to an immediate release/uptake of two electrons) are not taken into account in all these reaction cycles.

Both, the ligation sphere of molybdenum at the different reaction-steps and the mechanism of FDHs are still not completely clear. Therefore, quantum chemical studies, followed by Raman-spectra computations and comparison to experimental resonance Raman data (by Wahelfeld et al. [73]) have been carried out in this work. Theoretical models for these studies have been built and verified. Furthermore, different possible mechanistic features, like Cys-dissociation, SH-formation and formate-ligation, have been investigated in great detail. The determination of the complete ligation sphere of the Moco at different points in the catalytic cycle together with energy-computations will help to generate a new and more complete understanding of the catalytic oxidation of formate to carbon dioxide. This will help crossing the gap between current mechanistic understandings and provide helpful tools for future studies. The results of these studies are shown in part V.

Part II.

Materials and Methods

Chapter 4

Theoretical background

Theoretical/Computational methods are a necessary counterpoint to experiments. Many experimental observations cannot be interpreted completely, since information about events on the atomic and molecular scale is missing. Computational investigations provide helpful tools to understand experiments, such as measured spectra and to predict structures and reaction mechanisms. Additionally, they can shed light on dynamical properties and structural changes inside the investigated matter.

As stated in the sections before, there are many fields of application for theoretical studies. In this work dynamical properties of Cph1 have been investigated. Additionally, missing structural parameters of both Agp2 and *Rc*FDH have also been evaluated. Finally, for both latter systems mechanistic properties have been determined. For all these areas very different computational methods have been used, for which the theoretical background will be discussed in the following section. The focus will be mainly on the methods applied to above mentioned problems. While other methods may be available, they will only be mentioned.

4.1. Molecular Dynamics

Molecular dynamic (MD) simulations are a useful tool to investigate molecular systems in structural and dynamic terms at an atomic resolution. In these simulations, a trajectory will be generated over time, in which the time-evolution of velocities and positions of every atom and molecule is monitored in the context of its conformational space. Molecular dynamics are able to provide equilibrium geometries and transition states, as well as relative energies of conformers and potential energy surfaces. To get the most realistic expression for the investigated systems, the time-dependent Schrödinger equation [74] has to be evaluated for every atom at any given time-step. Since this is unfortunately only possible for very small systems, due to current computer power, different approaches are necessary to simulate different sizes of systems. MD-simulations are of special interest for the simulation of conformational space for liquids [75],[76],[77], periodic solids and surfaces [78],[79],[80], as well as biomolecular systems [81],[82],[83] (which are the area of interest in this work). These cannot be treated completely with quantum mechanics, since they are too large and too computationally costly. MD-simulations are the method of choice for these. Several methods exist and are prerequisite to carry out MD-simulations which will be discussed in the following.

4.1.1. Molecular Mechanics – Classical force fields

The simulation of large biomolecular systems, such as proteins including cofactors, *via* molecular dynamics is achieved by a simplificaton of matter. Classical force fields regard atoms as partially charged unflexible spheres. Classic Newton mechanics are applied to these. Covalent bonds between atoms are treated as classical springs, whereas electrons are completely omitted. These simplifications lead to the following analytical expression for the potential energy of the investigated molecular system as depicted in equation (eq.) (1) which is called a force field [84]:

$$V = V_{bonded} + V_{non-bonded} \qquad \textbf{(1)}$$

Here V is a function of positions r(N) of all nuclei.

Different terms attribute to the bonded interactions V_{bonded} [84], such as bond stretching V_{bond}, angle bending V_{angle}, bond rotation (torsion) $V_{torsion}$ and out of plane movements (improper torsion) $V_{improper}$.

$$V_{bonded} = V_{bond} + V_{angle} + V_{torsion} + V_{improper} \quad (2)$$

Further, non-bonded interactions $V_{non\text{-}bonded}$ are divided into electrostatic interactions (Coulomb potentials) $V_{electrostatic}$ and van der Waals (vdW) (London) interactions V_{vdW} [84],

$$V_{non-bonded} = V_{vdW} + V_{electrostatic} \quad (3)$$

Motions corresponding to bonds, angles and torsions, as well as vdW and electrostatic interactions are schematically illustrated in Figure 13.

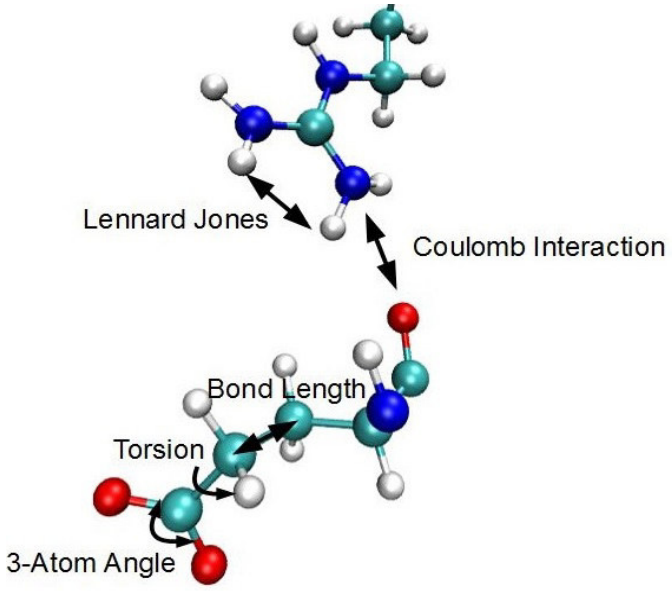

Figure 13. Schematic representation of different inter- and intramolecular motions (bonds, angles, torsions, vdW and electrostatic interactions).

One possible energy expression for the potential energy is the CHARMM(28) [85],[84] force field, as shown in eq. (4) which has been used for all MD-simulations carried out in this work:

$$V_{total} = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{torsions} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{improper} K_\phi (\phi - \phi_0)^2 + \sum_{non-bonded} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \quad (4)$$

Here $K_{(b,\theta,\chi,\Phi)}$ denotes the respective force constants for bonds (b), bond angles (θ), dihedral angles (χ), and improper dihedral angles (Φ), whereas b_0, θ_0, χ_0 and Φ_0 are the respective values at the equilibrium.

Both bond stretching and angle bending (as well as impropers) are given by a Taylor series around the equilibrium bond length b_0 (or equilibrium angle width θ_0 or width of improper Φ_0), following Hooke's Law. In both, the corresponding energy is assumed to increase quadratically (following a harmonic potential) with displacement from the equilibrium value. In contrast to this, the energy expression of the torsional motion is approximated by a Fourier series. In this way, the periodic nature of the torsion is captured. Some steric and electrostatic interactions between neighboring atoms can thus be mimicked (corresponding parameters are coupled to non-bonded ones, see next paragraph).

Electrostatic and vdW interactions (non-bonded) are described as a sum of Lennard-Jones and Coulomb terms. The first depends on the parameters ε_{ij} (depth of energy well) and R_{minij} (minimum energy distance), for the atom pair i and j and the distance r_{ij} between them. The Coulomb interactions are calculated between pairs of atoms at a relative distance r_{ij}, containing partial charges q_i and q_j and depending on the dielectric constant of vacuum ε_0.

The determination of the parameters define the accuracy and possible applications of a given force field [84]. It can be done in several ways, such as fitting to experimental values and explicit treatment of important parts of the system with quantum chemical methods. A given force field is limited to a finite number of systems, for which the parameters are derived for specifically [84]. Therefore, for different systems, different force fields have to be generated.

4.1.2. Molecular dynamics algorithms (Equations of motion)

The in equation (4) depicted expression for the potential is used to generate the force F(r(t) of any atom at each given time step as shown in (6). This is done by differentiating the potential as computed by the force field depicted in (4) over space. Numerically integrating Newton's equations of motion as shown in (5), in which F is the force acting on an atom and m its mass and a its acceleration then yields the molecular dynamics trajectory [86].

$$F = m \cdot a; \left(\frac{\partial^2 r}{\partial t^2} \right) = \frac{1}{m} F(r(t)) \qquad \textbf{(5)}$$

In the case of a 3N particle system this leads to a set of 3N coupled second-order differential equations of the form shown in (5) [86]. The force of one atom at position r(t) of the system is a function of all other atoms.

$$F(r(t)) = \frac{-\partial V(r)}{\partial r} \qquad \textbf{(6)}$$

There exist different algorithms for solving the differential equations described above. The so-called Velocity-Verlet algorithm [87],[86] is the one of choice in this work and the one most often applied in MD-simulations:

An initial guess for velocities v(0) at the beginning of the MD by a Boltzmann distribution [86] and starting coordinates r(0) (commonly obtained by crystallographic structures or NMR measurements, as well as homology models) are the prerequisites [86]. The timestep Δt has to be chosen one order of magnitude smaller than the fastest motion (which is the X-H stretching motion with 10^{-14}s) and thus be 1 fs [86]. To achieve a bigger Δt of 2 fs (and thus reduce the necessary integration steps)

further prerequisites have to be used, as explained in section 4.1.6. Finally, an appropriate expression for r(t) and v(t) at each given timestep for each atom is necessary for generation of a MD-trajectory [87],[86]:

In the framework of the Velocity-Verlet algorithm [87] these are described in equations (7) and (8):

$$r(t+\Delta t)=r(t)+\Delta t\cdot v(t)+\frac{\Delta t^2 F(r(t))}{2\mathrm{m}} \quad \textit{(7)}$$

$$v(t+\Delta t)=v(t)+\frac{\Delta t[F(r(t+\Delta t))+F(r(t))]}{2\mathrm{m}} \quad \textit{(8)}$$

The calculation of a molecular dynamics trajectory is then described in the following way, in which the necessary parts are computed for every atom at each time:

1. Starting point with r(0), v(0) and a(0).

2. Computation of r(t+Δt) using (7).

3. Evaluation of the potential V(r) as shown in (4).

4. With V(r) the force acting on the atoms F(r(t+Δt)) can be obtained using (6) and also the acceleration a(t+Δt).

5. Finally the new velocities v(t+Δt) are obtained according to (8).

This cycle is repeated until the initially defined number of timesteps is reached. The main advantage of this algorithm is that it is able to give positions and velocities at the same time [87].

4.1.3. Constant temperature/constant pressure dynamics

Biological systems (which are the subject of study in this work) mostly appear under temperatures of 300 K and a pressure of 1 atm. Therefore, it is important to maintain this conditions during the simulation. The volume is allowed to change during the simulation. Thus a so-called NPT ensemble is generated [86]. This means that the number of particles (N), pressure (P) and temperature (T) are kept constant and therefore the system is both isobaric and isothermic. To achieve this, a thermostat and a barostat are applied as described below [86]:

Constant temperature

The kinetic energy E_{kin} of a N-body system is coupled to its temperature [86] as given in (9), where N_A is the Avogadro number, K_B the Boltzmann constant and T(t) the temperature at any given time.

$$E_{kin}=\frac{2}{3}N_A K_B T(t)=\frac{1}{2}mv(t)^2 \quad \textit{(9)}$$

Since the kinetic energy is related to the velocity (eq. (9)), also v(t) and T(t) are proportional to each other and v(t) can be used for controlling the temperature of the system. This is done by scaling of

v(t) with a factor λ which is proportional to the temperature (both desired $T_{desired}$ = T_{bath} and the current one $T_{current}$ = T(t)) as shown in (10) [86]:

$$v_s(t) = \lambda \cdot v(t) \rightarrow \lambda \propto \sqrt{\frac{T_{desired}}{T_{current}}} \quad \textit{(10)}$$

This T-dependent factor can be generated by coupling an external heat bath (which has the desired temperature $T_{desired}$) to the system with the coupling parameter τ as shown by Berendsen [88],[89]:

In eq. (11) the corresponding expression for λ is given.

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right)} \quad \textit{(11)}$$

τ is defined as the step-size, in which the temperature is modified and adjusted from the current one to the desired one T_{bath} and should be chosen according to eq. (12), so that for a timestep of 1 fs a value of 0.4 ps is used.

$$\tau \rightarrow \frac{\Delta t}{\tau} \approx 0.0025 \rightarrow \Delta t = 1\ fs\ ;\ \tau = 0.4\ ps \quad \textit{(12)}$$

After T(t) equals the desired T_{bath}, it is constant at this value during the simulation.

Constant pressure

Maintenance of pressure can be achieved by changing of the volume of a given system [86]:

This is due to the fact that the volume is related to its isothermal compressibility κ, as shown in (13):

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \quad \textit{(13)}$$

A constant pressure can be obtained by scaling of the volume (represented by the positions of the atoms r(t)) with a factor λ [86]. In (14) the evaluation of the new volume r_s(t) via relation between λ and r(t) is depicted.

$$r_s(t) = \lambda^{\frac{1}{3}} r(t) \quad \textit{(14)}$$

λ can be gained by coupling of the system to an external pressure bath (15) [86],[88] (a commonly known way to couple a temperature bath to a pressure bath is the Nose-Hoover thermostat [87]).

$$\lambda = 1 - \kappa \frac{\Delta t}{\tau_P}\left(\frac{P(t)}{P_{bath}}\right) \quad \textit{(15)}$$

Here τ_p is the corresponding coupling constant which in a similar way as explained in eq. (12) represents the step-size, in which the current pressure P(t) is adjusted to the desired one P_{bath}. It is chosen according to (16).

$$\left(\frac{\partial P}{\partial t}\right) = \frac{P_{bath} - P(t)}{\tau_P} \qquad \textit{(16)}$$

4.1.4. Boundary conditions

Another important aspect of MD-simulations is the correct treatment of boundaries. In any system, the outer regions near the edges would feel different forces and interactions with surrounding atoms than the ones in the bulk. In native environments biological systems are much larger and also strongly interconnected. If only a single part of these macroscopic systems is to be simulated, it has to be truncated in its size (to reduce computational costs) and thus neglect the interactions of the atoms at the edges with potential neighbors. This might lead to wrong forces and wrong dynamics of the corresponding atoms. Therefore, a proper treatment of these areas has to be considered, so that every atom is treated as if residing in the center of the system. In the following the periodic boundary conditions [90] and stochastic boundary [91],[92],[93] concepts are described in more detail, since they have been used in this work.

Periodic boundary conditions (PBC)

PBC [90] can be used to reduce the number of artifacts, resulting from the interaction of the particles with the boundary region of the system: Translation operations are used to replicate the simulation cell in each direction. This generates a periodic system. The shape of the periodic cell can be e.g. cubic or hexagonal (as used in this work). Since in the image cells positions and velocities of the particles are not evaluated by solving differential equations (6) no computational costs do arise from these. In Figure 14a the general principle of PBC is shown.

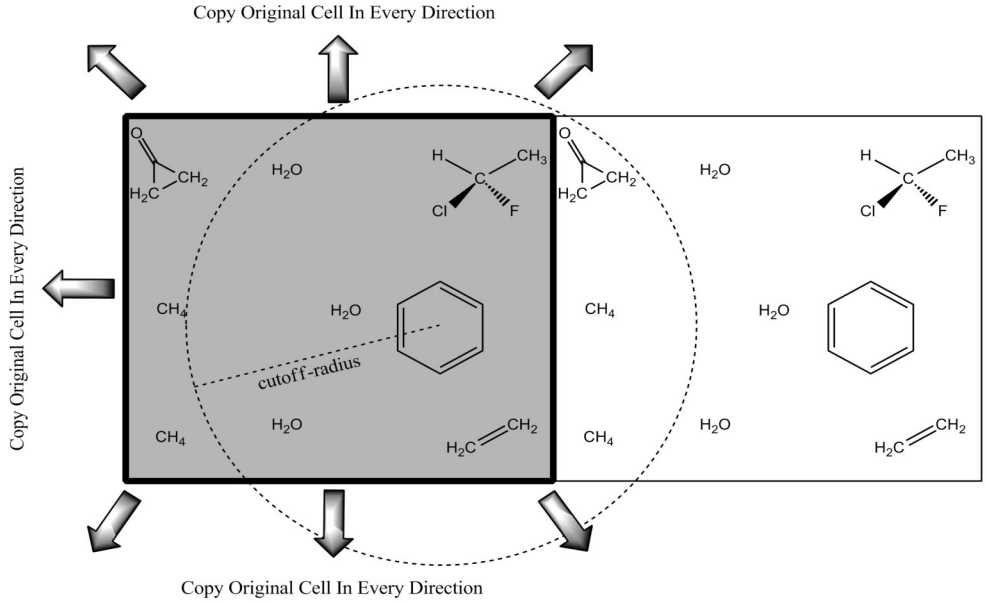

Figure 14a. Schematic representation of periodic boundary conditions. The cell in the middle is repeated in each direction in 3D space. A cutoff-radius is chosen for vdW interactions.

To guarantee that each atom sees only one version of the others and avoid its interaction with itself, a cutoff is chosen for the non-bonded interactions. This represents a radius, outside which all non-bonded interactions can be truncated and are not calculated. This follows the minimum image convention [94]. Such an approach reduces the necessary computational costs significantly. Additionally, the Particle Mesh Ewald [95],[96],[97] (PME) method is used to evaluate the electrostatic interactions in the chosen system with PBC. This is due to the fact that only fast-decaying interactions, such as van der Waals forces (decays with r^{-6}) can be treated effectively with truncation methods. Long-range interactions, such as coulombic ones (decaying only with 1/r) need to be treated differently, as shown by Ewald [95] who split 1/r into two fast-decaying terms. Simply spoken, in PME, the Coulomb interactions are computed by smearing the electronic charges of the system on a grid [96],[97].

Stochastic boundary conditions

There exist also other methods to treat boundaries, such as stochastic boundary conditions (SBC) [91],[92] and the generalized solvent boundary potential (GSBP) [93]. In both only the solvent in direct vicinity of the biological system is treated explicitly, while the solvent at higher distances is approximated by a dielectric constant or a repulsive force. While stochastic boundary conditions will be discussed in more detail, GSBP will not.

SBC is an alternative to PBC, in which the explicit treatment of solvent molecules is partly avoided [91],[92]: In many cases, only a small part of the investigated system is of interest. This can be for example an active center of an enzyme or a cofactor-binding region in proteins. Thus only a small amount of the protein has to be explicitly solvated, while the rest of the solvent can be treated in less costly ways, such as mimicking by a boundary potential F_b (r_i) which is felt by every atom at positions r_i [91],[92]. As depicted in Figure 14b, the system is spherically shaped and further divided into three parts (treated with different theoretical approaches), depending on the importance of the contained atoms [91],[92]: I reaction region, II boundary/bath region and III reservoir region:

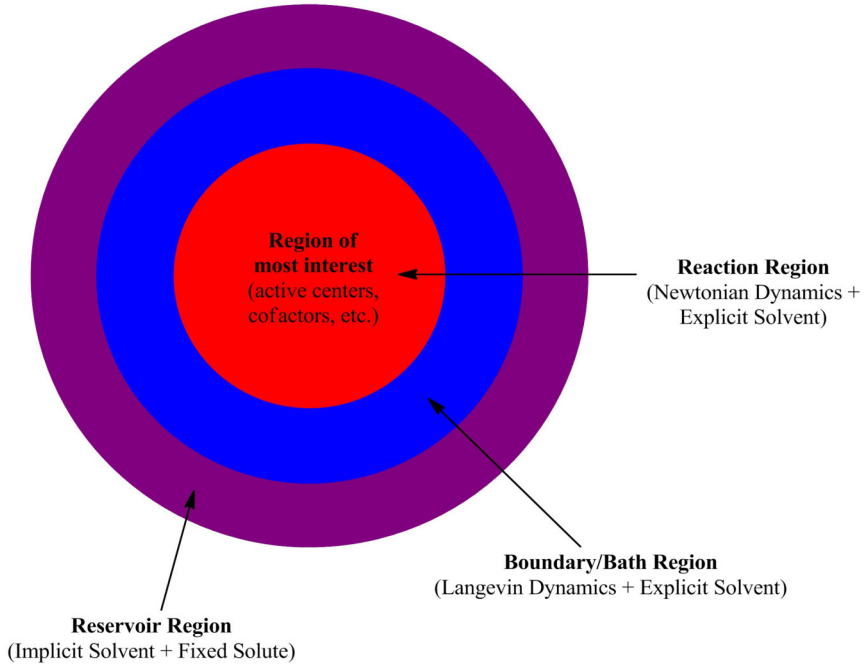

Figure 14b. Schematic representation of stochastic boundary conditions. The system is divided into three parts: (I) Reaction region, (II) Boundary/Bath region and (III) Revervoir region which are treated different with respect to levels of theory.

Reaction region

This region contains sites of interest, like active centers, cofactors and certain interesting residues/molecules. Therefore, all atoms (including solvent) are free to move and treated explicitly. The level of theory is depending on the type of computation carried out. In classical MD-simulations, all atoms in this region are treated according to Newton dynamics. To replace the pair-interactions with the solvent in the reservior region, the atoms in the reaction region are feeling an additional force $F_{add}(r_i)$ as depicted in eq. (17).

$$F_{add}(r_i) = F_b(r_i); r_i \leqslant r_{reaction} \quad \textbf{(17)}$$

In hybrid QM/MM calculations (s. section 4.3. for more detailed information about this method) this region is treated at a quantum chemical level.

Boundary region

This region serves as a buffer area between I and III, in which all atoms (solute and solvent) are treated explicitly according to Langevin dynamics. The atoms in this region are feeling an additional force $F_{add}(r_i)$ as depicted in (18).

$$F_{add}(r_i) = F_b(r_i) - \gamma_i v_i + R_i(t); r_{reaction} \leqslant r_i \leqslant r_{boundary} \quad \textbf{(18)}$$

Here, $\gamma_i v_i$ is the so-called frictional force (with γ_i the friction coefficient) which is attributing for the solvent-induced 'slow-down' the atoms are feeling. $R_i(t)$ is a random force driving the thermal motions of the atoms and represents the collision-effect solvent molecules in the reservoir region would have on particles in this area. While the frictional force is decelerating, the random force is stimulating the particles in this region. The latter is a Gaussian distribution centered at zero. Its width is defined by the temperature. The influence of these two additional forces is also visible in the Langevin-dynamics, as depicted in equation (19).

$$\frac{\partial^2 r_i(t)}{\partial t^2} = \frac{F_i(r_i(t))}{m_i} + \frac{R_i(t)}{m_i} - \gamma_i \frac{\partial r_i(t)}{\partial t} \quad \textbf{(19)}$$

Both $\gamma_i v_i$ and $R_i(t)$ are obeying to the fluctuation-disspiation theorem (20).

$$\int \langle R_i(t) \cdot R_i(0) \rangle \, dt = 6 K_B T \gamma_i \quad \textbf{(20)}$$

Here, K_B is the Boltzmann constant and T the temperature.

In this way the two additional forces are tuning the temperature of the boundary region. Thus this region serves either as a temperature sink or source for the energy in the reaction region and acts as a temperature bath. This preserves the equilibrium structure and reaction region in general for unnatural fluctuations.

Reservoir region

Solvent is not treated explicitly in this region. In classical MD-simulation, using stochastic boundary conditions, solvation effects are mimicked by a dielectric constant. A repulsive force $F_b^L(r_i)$ is contributing to F_{add} (eq. (21)) and is responsible for keeping the particles inside the inner sphere (reaction and boundary regions).

$$F_{add}(r_i) = F_b^L(r_i) - \gamma_i v_i + R_i(t);\ r_i \geqslant r_{boundary} \qquad \textbf{(21)}$$

STM are often applied to QM/MM computations. Thereby, part of the solute often reach into this region. All atoms herein are kept fixed. This is often according to the GSPB formalism [93] as a special case of stochastic boundary conditions which is not discussed here. Furthermore, no solvent is present outside a sphere of explicit water and the solute resides in vacuum. The force $F_b(r_i)$ is constituted by the partial charges of the solute. For further information about the used QM/MM methods in this work see sections 5.1.5 and 5.2.4.

4.1.5. Polarized force fields

In contrast to classical force fields, in which only point charges (derived from experiments) are used for investigated systems, a polarized force field provides a more realistic approach [67]. In the real system, the charges of molecules (e.g. cofactor) induce changes in the electrostatic field of its environment (e.g. amino-acids) and thus influence surrounding charges. In the same way, they are also under influence of the neighboring charge-environment. This charge-field is constantly changing, as the molecules and atoms move around in space. Polarization and depolarization effects are strongly changing over the course of the trajectory [67] (for more details and references on the importance of polarized force fields, please see sections 3.1. and 6. and 7.).

In a classic MD-simulation this is not taken into account: The charges remain constant for the whole simulation. Also the charges of the amino-acid residues are often derived without taking any possible present cofactors and their influence on the charges into account. One approach to achieve a more realistic description of electrostatics is to employ so-called polarized charges for at least the region of most interest (e.g. active center of an enzyme) [67]. These charges are represented by Mulliken charges [98] as discussed in section 4.2.4. The region of interest is simulated with QM methods, yielding more accurate charges due to the incorporation of electronic influences [67]. Afterward these new derived charges can then be incorporated into the classic CHARMM force field [85] (as shown in eq. (4)) and used for subsequent MD-simulations. Unfortunately, these charges cannot adapt to changes in the electrostatic environment during the course of the simulation. Nevertheless polarized charges represent an improvement compared to the classic MM approach, because influences of cofactors on amino-acids and *vice versa* can be accounted for [67].

4.1.6. Constrains and restrains of MD-simulations

In the case that e.g. some part of the system has not been parametrized properly or because of other reasons, it might be of importance to restrain or fix certain parts of the investigated system [99], [100]. In the latter case, the corresponding atoms remain fixed on their initial positions during the whole simulation by application of so-called constraints. One algorithm commonly used for internally constraining molecules is the SHAKE algorithm [86],[99]. In a similar way, X-H stretching motions can be constrained. This will lead to the possibility of an increased timestep of 2

fs and thus a reduction of computational costs and speeding up of the MD-simulation [86],[99].

Investigating biological matter often includes the solvation in huge amounts of water. To increase the speed of the simulation special potentials for water have been developed. In TIP3P [101] the O-H bonds, as well as the H-O-H angle are kept fixed to constant values. Thus water molecules can only move as rigid bodies. This potential for water (TIP3P [101]) is commonly used, when simulating biological systems.

Furthermore, it is possible to apply an additional force to an atom, such that it is less free in its movement. These is called restraining of atoms. Restrains are often used in pre-simulation adjustment of temperature to 300 and pressure as shown in section 4.1.3. Herein, the restraints are slowly reduced, until the corresponding atoms are completely free. This may be necessary to reduce the risk of sudden and unrealistic increases in the velocities (due to the increase in temperature) and thus the rise of artifacts.

Thus, constraints and restraints help to reduce computational costs, speed up the simulation, reduce artifacts and to overcome missing parameters.

4.1.6. Determination of structural properties by MD-observables

Several properties can be subsequently calculated from the evaluation of the positions of every atom at each time-step (MD-trajectory).

Radius of gyration (RGYR)

The RGYR is the divergence of an atom at $r_k(t)$ from the mean structure r_{mean}

$$RGYR = \frac{1}{N}\sum_{i=1}^{N}\left(r_k(t_i) - r_{mean}\right) \qquad \textit{(22)}$$

It is used for proteins to visualize and quantify the global movements of backbone atoms (i.e. folding or unfolding motions). It represents a “diameter” of the protein.

Root mean square deviation (RMSD)

The RMSD is another important property of the system which can be computed out of a MD-trajectory [102],[103],[104],[105],[106]. The basic principle is very similar to the one of the RGYR. It is the divergence of an atom at position $r_j(t_i)$ from a reference structure [102],[103],[104],[105], [106]. This can be the starting structure or a crystal model. Before the RMSD can be calculated, the molecular structure has to be aligned with the reference structure to exclude rotation- and translation-contributions.

$$RMSD(t_i) = \frac{1}{N}\sum_{i=1}^{N}\left(r_j(t_i) - \tilde{r}_j\right) \qquad \textit{(23)}$$

It can be used to determine, how much the computed structure deviates from the reference.

Root mean square fluctuations (RMSF)

It is computed again as a deviation between actual and reference structure [106]. The RMSF, in contrast to the RMSD, uses the structure averaged over all timesteps of the trajectory as the reference structure.

$$RMSF(j)=\sqrt{\frac{1}{T}\sum_{i=1}^{T}(r_j(t_i)-\langle r_j\rangle)^2} \qquad (24)$$

This value accounts for the flexibility of the corresponding parts of the system.

Further observables and properties of a MD-trajectory will be discussed in more detail in the results in parts III and IV of this work.

4.2. Quantum chemical methods

Quantum chemical methods provide a way to obtain a realistic description of the electronic structure of molecules. This is done in general by solving the Schrödinger equation following the Born-Oppenheimer approximation [107]. In this, nuclear-nuclear interactions are treated as constant, due to a separation of nuclear and electronic motions. This is valid, because electrons move faster than the atom cores and it is assumed that from the point of view of the electrons, the nuclei remain fixed. This leads to an electronic Schrödinger equation which can be solved in different ways. Either *ab initio* methods can be used, further divided into wavefunction-based methods and the ones based on the elctronic density, like DFT [108],[109],[110],[111],[112],[113], [114],[115],[116]. There exist also semiempirical methods, in which the above named methods are approximated with empirical data, so that computational costs can be saved. One example for the latter, is the SCC-DFTB method [117]. The application of these methods yield different properties of the system, like Mulliken charges and electronic density (in the case of DFT) and the appropriate wavefunction of the system. These will be further discussed in the following chapter. Since in this work only DFT-based methods, as well as SCCDFTB, have been used, no wavefunvtion-based techniques are discussed here.

4.2.1. Density Functional Theory (DFT)

The basic idea of the density functional theory is that each molecular system contains an electronic density [116]. This is calculated as shown in eq. (25) as the integral over all electronic coordinates in the investigated space [116]:

$$\rho(r)=N\int\ldots\int|\Psi(r,x_1,x_2,\ldots,x_N)|^2\,dx_1,dx_2,\ldots,dx_N \qquad (25)$$

Here Ψ is the wavefunction of the system and χ are the spatial spin-coordinates. ρ(r) is the electronic density which defines the probability to find any electron with exchangeable spin within the corresponding space, when all other electrons can be anywhere. Therefore, integration of ρ(r) will lead to the total number of electrons N of the system. And the density becomes zero, when r is approaching infinity (26).

$$\int \rho(r)\,dr = N\,;\, \rho(r \to \infty) = 0$$
(26)

The most important property of ρ(r) is that it is experimentally observable, in contrast to the wavefunction. This makes the use of DFT more intuitive and more easily comparable to experiments.

4.2.2. Hohenberg and Kohn

The development of the DFT method started with the proposals of the two fundamental theorems by Hohenberg and Kohn [108],[116]:

1. HK theorem:

Every observable quantity of a stationary quantum mechanical system is completely determined by the ground-state electron density ρ_0 (for example the ground-state total energy and the wavefunction).

The total electronic energy of the ground-state E_0 is defined as a sum of the kinetic energy T and the electron-electron interaction E_{ee} plus the electron-nuclear interaction E_{Ne} which all only depend on the electronic density (27):

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0]$$
$$E_0[\rho_0] = \int \rho_0(r) V_{Ne}\,dr + T[\rho_0] + E_{ee}[\rho_0]$$
(27)

In contrast to E_{Ne},, computed as an integral over the electronic density ρ_0 and the potential V_{Ne}, the kinetic energy of interacting electrons, as well as the electron-electron interactions are unknown [116]. These two energies form the so-called Hohenberg-Kohn functional F_{HK}, as stated in eq. (28).

$$F_{HK}[\rho_0] = T[\rho_0] + E_{ee}[\rho_0]$$
$$F_{HK}[\rho_0] = T[\rho_0] + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{r_{12}}\,dr_1\,dr_2 + E_{ncl}[\rho_0]$$
(28)

The electron-electron term is further divided into a classical Coulomb interaction E_H and a non-classical E_{ncl} (consisting of electronic exchange E_X and correlation E_C):

$$E_{ncl}[\rho_0] = E_x + E_c$$
(29)

If F_{HK} would be known exactly, the Schrödinger equation could be solved exactly for all molecular systems. This is due to the fact that F_{HK} is not unique for a given system, but a universal functional [116]. Finding F_{HK} is the major challenge of modern DFT-development.

2. HK theorem (varational principle)

The ground-state density can be calculated, in principle exactly, using the varational principle, involving only the density (Rayleigh-Ritz-principle) [109],[108],[116]:

The second theorem which is shown in equation (30) states that with the help of F_{HK} any given trial density would yield an energy which is higher than or equal to the exact ground-state energy. Provided that the correct density and F_{HK} of the true ground-state would be known, the exact energy could be calculated [108],[109],[116].

$$E_0 \leqslant E[\tilde{\rho}] = E_{Ne}[\tilde{\rho}] + F_{HK}[\tilde{\rho}] \quad \textit{(30)}$$

In summary it can be said that knowledge of the electronic kinetic energy of interacting electrons and the exchange and correlation energies would allow the determination of the correct ground-state and all its properties, including its energy (31).

$$E[\rho] = T[\rho] + E_{Ne}[\rho] + E_H[\rho] + E_x[\rho] + E_c[\rho] \quad \textit{(31)}$$

But since only E_{Ne} is known, a way to approximate or circumvent F_{HK} has to be found, done by Kohn and Sham [110].

4.2.3. Kohn and Sham

Since the kinetic energy of non-interacting electrons is known (32), Kohn and Sham introduced a fictitious non-interacting system that has the same density as the real one [110]:

$$T_{KS}[\rho] = -\frac{1}{2}\langle \phi_i | \nabla^2 | \phi_i \rangle \quad \textit{(32)}$$

Whereas Hohenberg and Kohn's approach has been orbital-free [108],[116], Kohn and Sham proposed an orbital-based approach, in which the orbitals Φ_i do not interact with each other [110], [116]:

A non-interacting density $\rho(r)$ (33) is generated as a linear combination of one-electron orbitals (33), named Kohn-Sham orbitals φ_i.

$$\rho(r) = \sum_{i=1}^{N} |\phi_i(r)|^2 \quad \textit{(33)}$$

Thus, the energy expression can be changed into (34) which is based on non-interacting electrons and orbitals:

$$E[\rho] = T_{KS}[\rho] + E_{Ne}[\rho] + E_{XC}[\rho] + (T[\rho] - T_{KS}[\rho]) \quad \textit{(34)}$$

Here T_{KS} is the kinetic energy of the non-interacting Kohn-Sham electrons and is dependent on the

Kohn-Sham orbitals φ_i. The unknown terms E_{XC} and the difference between the kinetic energy of interacting electrons and T_{KS} ($T(\rho)$-$T_{KS}(\rho)$) are further combined into the exchange-correlation functional (35).

$$E'_{XC}[\rho]=E_{XC}[\rho]+(T[\rho]-T_{KS}[\rho]) \tag{35}$$

This functional generates the so-called exchange-correlation potential V_{XC} (eq. (36)).

$$V_{XC}[\rho]=\frac{\partial E'_{XC}[\rho]}{\rho} \tag{36}$$

The one-electron energies ε_i, as well as the corresponding orbitals Φ_i are obtained by solving the one-electron Kohn-Sham equations, as stated in eq. (37) [110]:

$$\hat{f}_{KS}\phi_i(r)=\epsilon_i\phi_i(r) \tag{37}$$

Since starting values for φ_i are necessary to solve this set of one-electron equations as well as for defining the Kohn-Sham operator f_{KS}, they have to be solved iteratively and form a self-consistent field procedure. In this, an initial guess of the orbitals leads to an initial iteration with resulting energies and new orbital coefficients. These are then used for further iterations and calculations of f_{KS}, until a desired threshold is reached and the energy and coefficients converge. The used Kohn-Sham operator f_{KS} is defined thereby as [110]:

$$\hat{f}_{KS}=-\frac{1}{2}\nabla^2 V_{eff}(r);$$
$$V_{eff}(r)=\sum_{k=1}^{N_k}\frac{Z_k}{|r-R_k|}+\int\frac{\rho(r')}{|r-r'|}dr'+V_{XC}[\rho(r)] \tag{38}$$

The first term in V_{eff} displays the interaction of the electrons with the field induced by the nucelar charges Z_k. The second term describes the interactions of electrons with each other. The effective potential V_{eff} is strongly related to the exchange-correlation potential V_{XC}. This is further generated according to (36)

V_{eff} is important for the correctness of the chosen density and has be chosen in such a way that the non-interacting density equals the real (interacting) one. Therefore, a correct approximation for the unknown terms contributing to E'_{XC} has to be found. The gradient-corrected (like BP86) [111] and hybrid functionals [112] are the ones most widely applied. In the following work especially BP86 and the hybrid functional B3LYP (39) have been used.

$$B3LYP\rightarrow 0.8\mathrm{S}+0.72\mathrm{B88}+0.2\mathrm{HF}+0.19\mathrm{VWN}(III)+0.81\mathrm{LYP} \tag{39}$$

B3LYP consists of contributions from the Slater-Dirac-exchange term [113],[114], Becke-exchange

term, the Hartree-Fock-exchange term, the Voskov-Wilk-Nussair correlation [115] and the Lee-Yang-Parr correlation [111].

4.2.4. Mulliken charges

DFT computations can produce different properties of the investigated system, depending on its electronic structure. One of these are the charges of the individual parts of the system, the so-called Mullikan charges [98]. In the following the concept of Mulliken charges is discussed in further detail, since they have been used for the generation of polarized force fields (as discussed in section 4.1.5):

The Mulliken charges arise from the Mulliken Population analysis [98]. They can be used for the description of the electronic structure and also for determination of the bonding, antibonding or nonbonding nature of molecular orbitals (MOs). As prerequisites for the computation of Mulliken charges the molecular orbital wavefunction Ψ_i and the coefficients c_i of the corresponding atomic orbitals have to be known (generated for example with DFT methods as discussed in the previous section). According to the linear combination of atomic orbital (LCAO) approach [116] the charge distribution of the 2-particle system can be displayed as a probability density (40):

$$\Psi_i^2 = c_{ij}^2 \phi_j^2 + c_{ik}^2 \phi_k^2 + 2c_{ij} c_{ik} \phi_i \phi_j \qquad \textbf{(40)}$$

Here c_{ij} and c_{ik} are orbital coefficients, while Φ_j and Φ_k are the atomic orbitals (AOs). Integrating over all electronic coordinates, with normalized AOs and MOs, one gets the following expression:

$$1 = c_{ij}^2 + c_{ik}^2 + 2 c_{ij} c_{ik} S_{jk} \qquad \textbf{(41)}$$

Here S_{jk} is the overlap integral, describing the overlap between two AOs. This can be interpreted according to Mulliken [98] as follows:

c^2_{ij} is the contribution of one electron in MO Ψ_i to the electronic charge in AO Φ_j. Accordingly c^2_{ik} is the contribution of one electron in MO Ψ_i to the electronic charge in AO Φ_k. Hence these are called atomic-orbital populations. Furthermore, $2c_{ij}c_{ik}S_{jk}$ is the contribution of one e^- to the overlap region between the two corresponding AOs and is called overlap population (>0: bonding MO, <0: antibonding MO, =0: nonbonding MO). These results are commonly presented in the so-called Mulliken population matrix:

$$P_i = \begin{pmatrix} c_{ij}^2 & 2c_{ij} c_{ik} S_{jk} \\ 2c_{ij} c_{ik} S_{jk} & c_{ik}^2 \end{pmatrix} \qquad \textbf{(42)}$$

Here rows and columns correspond to atomic orbitals, while each diagonal element displays the atomic-orbital populations with the off-diagonal elements being the overlap populations.

Further condensation of the data can be achieved *via* combination of the atomic-orbital populations with the correlated overlap populations for each MO which result in the Gross population matrix [98]. The matrix elements are created by dividing the overlap populations in half and adding each half to the AO populations of the participating Aos (43):

$$GP_{ij} = P_{i_{jj}} + \frac{1}{2}\sum_{k \neq j} S_{jk}$$
(43)

The GP matrix and its elements provide the amount of gross charges that the MOs contribute to the AOs. Columns correspond to MOs, while rows correspond to AOs. Furthermore, AO charges Q_φ are then obtained by adding the elements in the rows of GP for the occupied Mos.

Atomic charges Q_{atom} are obtained from the Q_φ when adding the AO charges on the same atom according to eq. (44).

$$Q_{atom} = \sum Q_{\varphi_k}$$
(44)

And finally the Mulliken charges $Q_{Mullikan}$ (also called polarized charges) are obtained by subtracting Q_{atom} from Z_{atom} (charge of the nucleus) according to (45):

$$Q_{Mullikan} = Z_{atom} - Q_{atom}$$
(45)

4.2.5. Basis functions and basis set approximations for metals

An important prerequisite for *ab-initio* calculations is the correct definition of the atomic orbitals. As shown before, one-electron orbitals are used in DFT [110] for iterative computation procedures. In the MO-theory, the LCAO approach defines the description of molecular orbitals by a linear combination of atomic orbitals [116]. To achieve reliable results, a good approximation for these atom-centered orbitals has to be found. In the following the theory of basis sets is explained, as well as different kinds of basis sets and their possible application [116],[118]. Finally, basis set approximations for d-period metals are discussed, since computations involving metals have been done in this work [119],[120],[121],[122].

Standard Basis functions

Atomic orbitals are represented in QM calculations by so-called basis functions [116],[118]:

In general, basis functions are predefined and reflect important properties of the atoms they should resemble. They are centered on the nucleus and are similar to the corresponding atomic orbitals. Commonly used are the so-called gaussian type orbitals (GTO) as shown in eq. (46)

$$GTO: R(r) = r^l e^{-\xi r^2}; STO: R(r) = r^{n-l-1} e^{-\xi r}$$
(46)

Here r and l represent the radius of the orbital and the angular momentum quantum number, respectively. ξ is related to the nuclear charge of the corresponding atom. Although functions of this type are easy to compute, they are a bad basis for actual AOs, since they differ much from the exact solution of the Schrödinger equation, as shown in eq. (47) [116].

$$R(r) = L_{n-l-1}^{2l+1}\left(\frac{2Zr}{n}\right)e^{-\frac{z}{n}r} \qquad (47)$$

L corresponds to the radial part of the atomic orbitals and is a Laguerre polynomal. It is dependent on the different quantum numbers n (main quantum number) and l, as well as the nuclear charge Z and the radius r.

Slater type orbitals (STOs) are realistic representations for AOs (see eq. (46); not discussed here), but are computationally very costly in contrast to GTOs.

Therefore, a linear combination of several GTOs has to be used to approximate the radial distribution of the atomic orbitals. A number of primitive GTOs (PGTO) is forming contracted GTOs (CGTOs) which are shown in eq. (48).

$$\chi(CGTO) = \sum_{i}^{k} a_i \chi_i(PGTO) \simeq STO \qquad (48)$$

This provides a good approximation for real atomic orbitals and is similar in quality to STO (if still less flexible) and less computationally costly. Different basis set types are vailable [116],[118]:

Basis set. A set of CGTOs then form a basis set, used for description of AOs/MOs. The appropriate coefficients have to be determined e.g. by DFT computations. There exist different basis sets which differ in its application and accuracy:

Minimal basis. In a minimal basis a single CGTO is used to describe every orbital (e.g. STO-kG, where k denotes the number of PGTOs to be contracted for each atomic orbital).

Valence double zeta (VDZ) basis. VDZ basis sets apply two CGTOs to each orbital. Triple zeta basis sets use accordingly three CGTOs. More commonly used are the so-called split valence basis sets, in which the core orbitals are represented in minimal basis, while the valence orbitals are represented by VDZ or VTZ basis sets. This is due to the fact that the core electrons/orbitals are mostly unimportant for chemical behavior, while the valence orbitals contribute heavily to binding situations in molecules.

Polarization functions. Polarization functions can be applied to achieve a better description of bonding situations. They add a function with higher angular momentum (e.g. for hydrogen: three p-functions are added; for carbon: five d-functions are added).

Diffuse functions. These functions have small exponents and are added to the basis set for the purpose of better describing diffuse states, such as anions and hydrogen bridges (weak bonds in general).

Effective core potential (ECP) → pseudopotential

Another important approximation for basis sets is the so-called effective core potential (ECP) or pseudopotential [119],[120],[121]. Because of the low contribution of core-near electrons to

chemical reactions and their orthogonality to the valence electrons, the all-electron potential can be approximated by an appropriate potential, named ECP. The valence electrons move inside this potential. It was first introduced by Hans Hellmann [119]:

Using ECP's, the solution of the Schrödinger equation becomes much less costly in terms of computational resources. There are three different approximations for generation of ECP's:

1. One-electron approximation:

Separation of core and valence electrons.

2. Frozen core approximation:

The one-electron contributions of the core-near orbitals are constant.

3. Small-core approximation:

In the small-core approximation it is assumed that there is no significant overlap between core and valence electrons which is why the former can be kept constant. For cases, in which the overlap between both is not negligible their exist other methods like non-linear core corrections [120] and semicore electron inclusions [121].

This leads to a simplification of the DFT approach, in which the exchange-correlation functional can be separated into two terms (49). Thus, in DFT computations using ECPs only the part for valence electrons has to be considered.

$$E'_{XC}(n_{core}+n_{valence}) = E'_{XC}(n_{core}) + E'_{XC}(n_{valence}) \quad \textbf{(49)}$$

Thus, the valence electronic-potential in ECPs is described by so-called pseudo-wavefunctions, while the core-electrons are omitted. After a chosen cut-off, the pseudo-wavefunctions and the real potential function are similar. A bigger cut-off guarantees a better convergence, but makes the ECP less transferable to different systems.

Incorporation of ECP's reduce not only the necessary basis set size and the number of electrons (and thus the computational costs), it also addresses heavy metals [122]: Relativistic effects can be found in very heavy elements, like d-row-metals: Because of the heavy core (and higher positive charge), the core electrons are moving at a much higher velocity, nearly with the speed of light, resulting in the contraction of the core orbitals and therefore, stronger shielding of the valence electrons from the core charge. That is why, the valence orbitals are more spacious than normally assumed and are also much more separated from the core orbitals.

Therefore, in d-metals, core-electrons play a less significant role in molecular-binding situations. Further the contribution of valence orbitals to chemical bonds is altered compared to non-relativistic elements. This cannot be approximated realistically by standard basis sets. Because of this, the application of ECP's for d-metals is of high importance to simulate correct electronic behavior and get realistic results [122]. QM-computations including d-metals, as have been done in this work, are therefore always treated with appropriate ECP's and basis sets (for further detail about the used basis sets, please see section 5.2.).

4.2.6. Self-Consistent Charge Density Functional Tight Binding method (SCC-DFTB)

The SCC-DFTB method [117] is an approximation of density functional theory (with use of the Kohn-Sham orbitals [110]): It uses empirical parameters and is a semiempirical method. Further it only takes valence electrons into account, therefore using pseudopotentials. This leads to computational efficiency at low costs (compared to conventional DFT).

The energy function is divided into three different parts (50):

$$E_{tot} = E^{H0} + E^{\gamma} + E^{Rep};$$
$$E_{tot} = \sum_{i\mu\nu} c_{i\mu} c_{i\nu} H^0_{\mu\nu} + \frac{1}{2} \sum_{A,B} \Delta q_A \Delta q_B \gamma_{AB} + E^{Rep} \quad \textbf{(50)}$$

The hamiltonian energy E^{H0} is generated by using either analytical functions or pre-evaluated results from DFT calculations (minimal basis of pseudoatomic orbitals with usage of GGA functional) for the desired system or molecule. These have been carried out using a reference density which contributes to the Tight-binding (TB) model (atom hardness, as is typical for ECP-treated orbitals). As seen in (50), the first term in E^{tot} is defined as the sum over the product of the atomic coefficients c_i in minimal basis representation and the Hamilton H^0 matrix, depending on the reference density and V_{XC}.

The second term E^{γ} is strongly related to the exchange-correlation functional. Δq is the induced charge on the corresponding atom and γ_{AB} a distance-dependent function which describes charge interactions and approximates E'_{XC} with a Taylor series (self-consistent charge method is applied) to account for changes in charge-densities and charge transfers. In SCC-DFTB first and second order terms are used for this, resulting in a good description of short- to long-range charge interactions.

The last term, E^{Rep}, the repulsive part, is built from experimental parameters resulting in the approximation of the two-center integrals by one-center terms. This is done by fitting parameters to experimental values. Further E^{Rep} is directly related to the application of pseudopotentials and thus the neglect of core electrons.

The coefficients and energies are determined by solving the Kohn-Sham equations as shown in section 4.2.3., resulting in equations shown in (51) which are solved iteratively. $H^0_{\mu\nu}$ is preevaluated and similar to the Kohn-Sham operator in standard DFT (eq. (38)) and depends on the reference density ρ^0.

$$\sum_{\nu} c_{i\nu} (H_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0;$$
$$H_{\mu\nu} = \langle \phi_\mu | \hat{H}^0_{\mu\nu} | \phi_\nu \rangle + \frac{1}{2} S_{\mu\nu} \sum_C \Delta_{qC} (\gamma_{AC} + \gamma_{BC}) \quad \textbf{(51)}$$

These simplifications yield a semiempirical method on one hand more efficient than conventional MM methods and on the other hand less computationally costly than standard DFT [110] and therefore, a convenient choice for e.g. QM/MM hybrid approaches (as discussed in 5.1.3.).

4.2.7. Implicit solvation methods – Self-Consistent Reaction Field (SCRF)

In this work theoretical methods are applied to biological systems and therefore, solvation effects on the biological matter have to be accounted for. As mentioned in section 4.1., the water environment can be treated explicitly or implicitly. The application of large amounts of explicit solvent molecules in DFT computations is unfavorable due to the limitations of current computer power. Therefore, the application of implicit solvation methods [123],[124],[125],[126],[127],[128],[129], [130],[131],[132],[133],[134],[135],[136] is of great importance and interest. Not only water environments can be simulated by implicit solvation models, but also the protein environment which surrounds e.g. cofactors can be accounted for with implicit solvation [71],[137],[138],[139].

Implicit solvation methods as applied in this work use a polarizable continuum (dielectrics) for mimicking the solvent. The first descriptions of molecules surrounded by a dielectric medium with incorporated charges can be seen in the classical Poisson-Boltzmann equation [126] and the generalized Born formalism [127],[128],[129].

A dielectric medium [140] generally has a high polarizability and is based on its dielectric constant ε which defines the permittivity of the material. Permittivity accounts for how much an external electric field has an effect on the material (and *vice versa*). The dielectric constant ε is the product of the relative permittivity ε_r (relation of ε to ε_0) and the one in vacuum ε_0 as depicted in (52). High values of ε_r account for high polarizability.

$$\epsilon = \epsilon_r \cdot \epsilon_0 \qquad \textbf{(52)}$$

In SCFR-calculations, the solute is surrounded by a dielectric continuum, defined by its dielectric constant and thus its polarizability. One way of approaching the computation of such models is shown in more detail in the next subsection. This method has been used in this work to mimic both solvents and protein environments.

Polarizable Continuum Models (PCM)

The polarizable continuum model (PCM) is the most common used SCRF method and has been derived by Tomasi et al. [123],[124],[125],[130],[131],[132],[133],[134],[135],[136],[141]: A cavity is created inside the dielectric continuum, in which the solute (molecule) is placed (as can be seen in Figure 15). This is carried out by using interlocking spheres defined by the van der Waals-radii *a* of each atom.

The solvent-solvent interactions are reduced upon introduction of the cavity inside the solvent resulting in a reduced entropy. Therefore, the energy ΔG_{cav} is needed to create the cavity which is therefore energetically unfavorable. In contrast to this, placing of the solute inside this cavity is energetically favorable, because of increased dispersion and electrostatic interactions, yielding ΔG_{disp} and ΔG_{ele} (electrostatic interaction). This leads to eq. (53).

$$\Delta G_{solv} = \Delta G_{ele} + \Delta G_{cav} + \Delta G_{dis} \qquad \textbf{(53)}$$

The individual terms are:

ΔG_{ele}: electrostatic solute-solvent interactions

ΔG_{cav}: energy needed to form the molecular cavity inside the continuum

ΔG_{dis}: dispersion energy arising from solute-solvent interactions

Additional terms may be (these are not always accounted for in available methods and are therefore not further discussed here):

ΔG_{rep}: repulsion energy (arising from Pauli repulsion)

ΔG_{Mm}: entropic contributions (due to molecular motions)

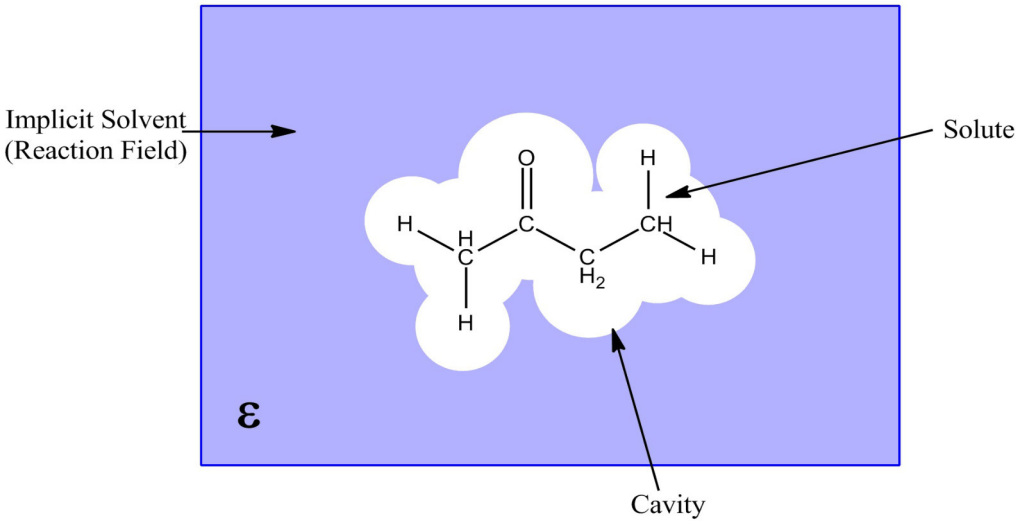

Figure 15. Polarizable continuum model: A cavity is build around the molecule/the solute, which is surrounded by the solvent, represented by a dielectric medium/reaction field. The picture was inspired by [135].

As said before, the cavity can be either be a simple sphere (radius a) with a point charge q in its center. Or it can be constructed in a more sophisticated way by overlapping spheres centered on each atom and defined by their van der Waals-radii.

For computing the individual terms of G_{solv} (as shown in (53)) three slightly different versions of cavities are used (e.g. [142]):

G_{cav} is calculated using the van der Waals-surface, generated by the van der Waals-radii of the individual atoms. The surface available for the solvent is then described by a so-called Solvent Assessable Surface (SAS) which is used to compute G_{dis}. The SAS is generated by inclusion of idealized solvent-radii into the van der Waals-surface. Finally, the G_{ele} is computed, using a third version of surface, namely the Solvent Excluding Surface (SES): Here, all van der Waals-radii are scaled by a constant factor and smoothed by the additional usage of not atom-centered spheres.

The energy used for creation of the cavity. ΔG_{cav} as well as the gained dispersion energy ΔG_{dis} is further proportional to the cavity's surface S, as shown in eq. (49) [135]. ξ is the proportionality factor.

$$\Delta G_{cav} + \Delta G_{dis} = \sum_{i}^{atoms} \xi_i S_i$$

(54)

The electrostatic energy is dependent on the polarization effect of solvent on solute and *vice versa* [135]: The induced dipole moment μ of the solute leads to the creation of charges on the molecular surface (Apparent Surface Charge (ASC) model). This forces the solvent to react/adapt to it which is why it is called a reaction field. Following this, the polarization of the solvent also induces a change in the molecular charge-distribution (multipole). In this way the solvent is also influencing the solute. The amount of the multipole moment change is determined by the polarizability α of the molecule. This solute-solvent interaction is defined in the so-called interaction potential V_{inter} which is added to the conventional Hamiltonian and used in the SCF procedure, as described below.

In the following the implementation of PCM into SCF procedures, as well as the analytic form of G_{ele} is illustrated for a simple spherical cavity [135]**:**

Eq. (55) [135] displays the analytic form of G_{ele}. Here, μ is the dipole moment of the solute and α its polarizability. ε_r is the relative dielectric constant of the dielectric medium surrounding the cavity and *a* the radius of the latter. The first part (before the brackets) corresponds to the polarization of the solvent and the second part (inside the brackets) to the back-polarization of the solute [135]:

$$\Delta G_{ele} = -\frac{\mu^2}{a^3}\frac{\epsilon_r - 1}{2\epsilon_r + 1}\left[1 - \frac{\epsilon_r - 1}{2\epsilon_r + 1}\frac{2\alpha}{a^3}\right]^{-1} \propto V_{inter}$$

(55)

The electrostatic free energy and the induced dipoles are directly proportional to the interaction potential V_{inter} which describes hence the interaction between solute and solvent. It is an additional term acting on the Hamilton operator in the Schrödinger equation as seen in eq. (56).

$$[H_0 + V_{inter}]|\Psi\rangle = E|\Psi\rangle$$

(56)

The SCRF model is then implemented into QM (DFT) calculations in the following way [135]:

1. Usage of normal SCF procedure (e.g. DFT) with conventional Hamiltonian H_0 (e.g. as discussed in section 4.2.3.). This results in the calculation of wavefunction $\Psi_0(r)$ and electronic density $\rho_0(r)$ in vacuum.

2. Calculation of dipole moment μ of the solute with $\Psi^0(r)$ according to (57):

$$\vec{\mu} = -\int \rho_0(r)\, dr + \sum_a Z_a R_a$$

(57)

Here, Z_α and R_α are the charge and coordinates of the nuclei, respectively.

3. The obtained dipole moment μ is then used to compute the induced reaction field E_R (58).:

$$\vec{E}_R = -\frac{2(\epsilon_r - 1)}{(2\epsilon_r + 1)a^3}\vec{\mu} \qquad \textbf{(58)}$$

4. This leads to the interaction potential V_{inter} (59):

$$\hat{V_{inter}} = \hat{\mu} \cdot \vec{E}_R \qquad \textbf{(59)}$$

5. Finally, the wavefunction $\Psi(r)$ and density $\rho(r)$ can be computed including the solvation-effects by adding V_{inter} to H_0, resulting in the new Hamiltonian H for the solvated system (60):

$$H = H_0 + V_{inter} \qquad \textbf{(60)}$$

6. The cycle can start again with step 2 and is repeated until μ, ρ and E_R converge.

The general procedure of first computing $\Psi_0(r)$ and afterward $\Psi(r)$ with the help of V_{inter} can also be applied to computations with non-spherical cavities.

This method is easy to implement and has lower computational costs as explicit solvation. Unfortunately there are some drawbacks to this model [130],[131],[132],[133],[134],[135],[136], [141]: First, ionic solutes which interact strongly with the solvent are not accounted for. Second, systems which include strong electrostatic interactions are less accurately treated, yielding often false positive charges in the system (especially if the wavefunction penetrates the cavity walls). Third, the estimation of non-polar entropic interactions (ΔG_{Mm}) are often neglected or treated insufficiently. And finally any interaction between solvent molecules and the solute (like H-bridges) are omitted. These can lead to wrong estimations of solvation energies and therefore, models with at least some of these qualities have to be treated with care, when applying PCM's.

4.3. Quantum Chemical/Molecular Mechanical (QM/MM) methods

Both pure MM and QM methods have huge advantages, but also shortcomings. MM is very useful for descriptions of conformational changes in big macromolecules and for analyzing dynamics. Unfortunately MM techniques are not able to describe the structure of the investigated systems on an electronic level. This means it cannot attribute for complete understanding of reaction mechanisms, charge transfers and electron excitations. In contrast, solving the Schrödinger equation leads to an understanding and visualization of the electronic structure. But since the computational costs for these methods are very high, they can only be applied to small systems and not to big biological systems, such as proteins. Clearly both methods have huge advantages and to overcome their restrictions Warshel and Levitt [84],[143],[144] introduced the concept of combining quantum mechanics with molecular mechanics in the QM/MM framework:

QM/MM is a hybrid method combining the powers of both MM and QM in one new concept. In this approach [84],[143],[144], the entire system (S) is partitioned into three different areas as depicted in Figure 16:

The outer region (O) described by molecular mechanics methods. According to stochastic boundary conditions [91],[92] (s. section 4.1.4.) this region can be further divided into Newtonian, Langevin and Boundary regions or at least be divided into an active and an inactive region, in which atoms are free to move or remain fixed, respectively. All atoms in the Newtonian area are completely unrestraint in their movements and follow Newtons laws of motion. In contrast to this, all atoms inside the second section experience additional frictional and random forces according to Langevin dynamics. Solute reaching out of the boundary region is treated as fixed (similar to GSPB [93]). Furthermore, most often the solvent is cut down to a sphere of a predefined diameter. Outside of this sphere solvent can be mimicked by application of a dielectric constant or completely omitted. See sections 5.1.5 and 5.2.4. for more information on practical application of QM/MM methods on biological systems, as carried out in this work.

The inner region (I) is described by QM methods. This is the area of most interest (e.g. active centers of proteins, reaction centers and so on). In the additive scheme both (I) and (O) are connected with a so-called boundary region (B). A general scheme of the QM/MM approach can be seen in Figure 16.

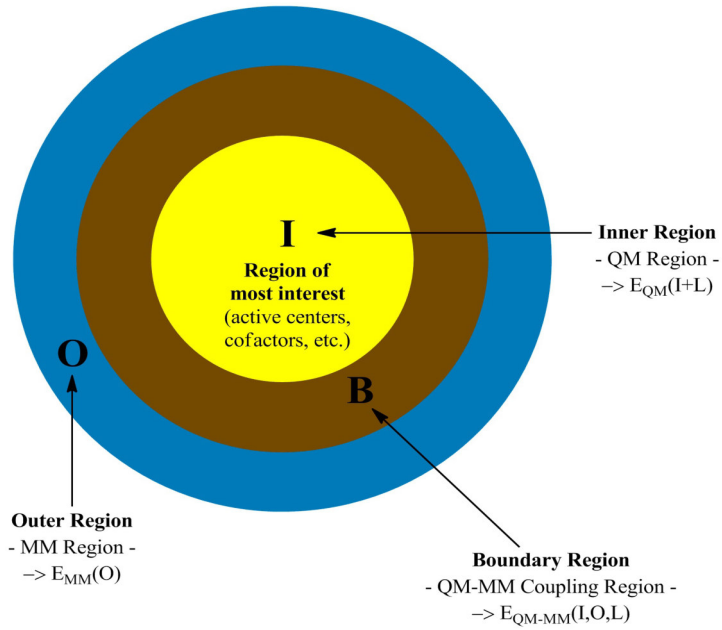

Figure 16. Schematic representation of a QM/MM model in the additive scheme as described in the text. The outer region (blue; O) is treated with MM, while the inner region (yellow; I) with QM methods. The brown area dividing both is called boundary region, concerning the transition from I to O and can be treated at different levels of theory.

Since both regions are strongly interconnected and interact with each other (e.g. *via* covalent bonds, which cross from (I) to (O) or with simple electrostatic interactions between both), it is not possible to simple sum up the energies of both regions to gain the one of the complete system (S) (61):

$$E(S) \neq E(I) + E(O) \qquad \textbf{(61)}$$

Instead, one has to think about coupling terms for both regions and take special precautions for the boundary region. This will lead to a new expression of the energy of S as shown in 4.3.1.

4.3.1. Additive Scheme (+ subtractive scheme)

Currently two different schemes for generating the energy of (S) are in use [84],[144]: One is called subtractive scheme and the other additive scheme.

Subtractive scheme

Here [84],[144], the QM/MM energy is generated (according to eq. (62)) by first calculating the energy of the complete system (S) with MM methods. Afterward, the MM-energy of (I) $E_{MM}(I+L)$ (with possible link atoms (L), as described in 4.3.3.) is subtracted from $E_{MM}(S)$ and the QM-derived energy $E_{QM}(I+L)$ of this region is added.

$$E_{QM/MM}(S) = E_{MM}(S) - E_{MM}(I+L) + E_{QM}(I+L) \quad \textbf{(62)}$$

Fortunately, this approach does not need any coupling terms for the QM and MM regions, since the complete system is initially computed with MM-methods. Second, the corresponding QM and MM packages do not have to be altered and thus the approach is easy to apply to existing software. Unfortunately, the coupling between QM and MM regions is only done with MM methods (leaving out electronic effects) and force field parameters are necessary for (I). This often proofs to be difficult, since appropriate parameters for e.g. cofactors might not be available.

Additive scheme

The additive scheme is targeting these disadvantages in the following manner [84],[144]: MM methods are only applied to the outer region (O), while the inner region (I+L) is treated at a quantum-chemical level. Both energy terms are combined. Furthermore, a third term is added which is called the QM-MM coupling term $E_{QM\text{-}MM}(I, O, L)$ and corresponds to the boundary region (B):

$$E_{QM/MM}(S) = E_{MM}(O) + E_{QM}(I+L) + E_{QM-MM}(I, O, L) \quad \textbf{(63)}$$

The coupling term treats interactions between the QM and MM regions ((I) and (O), respectively), such as bonding (covalent bonds across the boundary): E_{QM-MM}^{bond} , van der Waals interactions: E_{QM-MM}^{vdW} and electrostatic interactions: E_{QM-MM}^{elec} .

$$E_{QM-MM}(I, O, L) = E_{QM-MM}^{bond} + E_{QM-MM}^{vdW} + E_{QM-MM}^{elec} \quad \textbf{(64)}$$

This approach is the most used one in the field of QM/MM. The treatment of $E_{QM\text{-}MM}(I, O, L)$ and the corresponding individual interaction energies is therefore of major interest and can be approached in different ways [84],[144]:

E_{QM-MM}^{vdW} is commonly treated at the MM level by application of a Lennard-Jones potential. This makes it necessary to have correct van der Waals-parameters for both regions (I) and (O).

Covalent bonds, attributing to E_{QM-MM}^{bond} can be treated in different ways, as shown in 4.3.3 and the different possibilities for treatment of the electrostatic coupling between QM and MM atoms are discussed in the next section.

4.3.2. Embedding methods

The E_{QM-MM}^{elec} term can be addressed in different ways, which are called "embedding" methods [84],[144], attributing for possible polarization effects between both regions (I) and (O):

Mechanical embedding

The simplest way is to treat the QM-MM electrostatic interactions on a pure MM level. Point charges as common for force fields are also applied to the QM atoms. This leads to an exclusion of polarization effects of both region (I) and (O), since no charges on either side are polarized by the other. Also it is necessary to have appropriate point charges for the atoms of (I).

Electrostatic embedding

In this approach, the Hamiltonian of the the QM-region (I) includes the MM point charges of (O), as seen in eq. (65):

$$H_{QM-MM}^{\hat{elec}} = -\sum_{i}^{N}\sum_{j\in O}^{L}\frac{q_j}{|r_i - R_j|} + \sum_{a\in I}^{M}\sum_{j\in O}^{L}\frac{q_j Q_a}{|R_a - R_j|} \qquad \textbf{(65)}$$

Here q_j denotes the charges of MM atoms, while the charges of the QM atoms are Q_a and r_i and R_a the corresponding positions of the charges.

The incorporation of the MM charges leads to a polarization of the charges of the atoms of (I) and in this way (I) is polarized by (O). Therefore, this method is an improvement to the mechanical embedding, although it still lacks the back-polarization effect on MM atoms by the QM charges.

Polarized Embedding

The polarization of MM atoms in (O) by QM atoms in (I) is accounted for by polarized embedding methods. They work similarly to the electrostatic embedding, with the exception of using polarizable MM force fields. Flexible MM charge models are able to yield adjustments of the charges of MM atoms to neighboring charges (both QM and MM atoms) at each time-step of the computation. An example is the Drude-oscillator [84].

This method would provide the most accurate description in the QM/MM framework, but since polarizable force fields are still in development, the electrostatic embedding has been used in this work.

4.3.3. Boundary Region

The treatment of covalent bonds, which are cut by the QM-MM boundary, is also of great importance. They have to be included into the QM-MM coupling term according to equation (64).

Different approaches for generation of E_{QM-MM}^{bond} are available [84],[143],[144],[145],[146],[147], [148],[149] as discussed below.

Frozen orbitals approach

The first approach is the so-called frozen orbital or localized orbital approach. In this, frozen hybrid orbitals are used to saturate the free valency of the atoms at the cut bond [84],[143],[145],[146], [147].

Link-atom method

The second method (which has been applied to QM/MM computations carried out in this work) is the link-atom scheme which introduces an additional atom (not part of either MM or QM regions) connected to the QM host at the cut covalent bond [84],[144],[148],[149]: This atom is often a simple hydrogen atom H which is commonly included into the QM Hamiltonian and thus saturates the QM-atom valency. The link-atom interacts with the MM region only via electrostatics. In Figure 17 the basic principle of this approach is shown.

The MM-atom sees and interacts with the QM-atom on a classical MM-level via harmonic bond stretching as discussed in section 4.1.1. Since the MM-atom is very near to the H, compared to the original bond between QM and MM host atoms, this would lead to an overpolarization of the QM-charge in electrostatic and polarized embedding. Therefore, it is common to smear the charge of the MM atom on the neighboring atoms and thus "erase" it from the calculations.

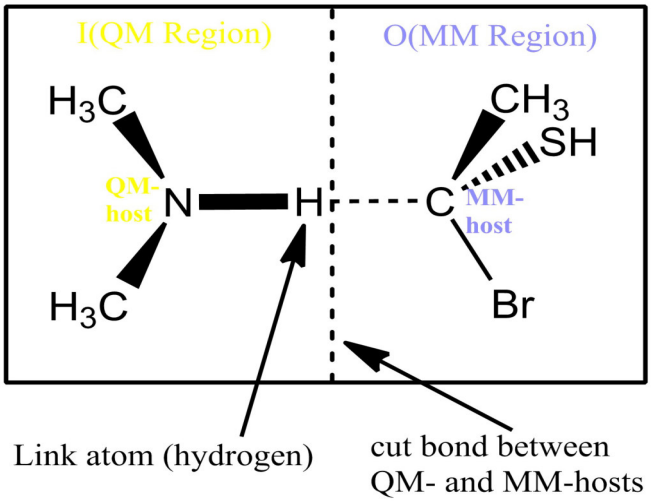

Figure 17. Schematic visualization of the link-atom method. The bond from the QM-host to the MM-host is cut. The QM-atom is then saturated by a hydrogen atom, while the MM-atom still sees the QM-host on a MM-level of theory.

4.3.4. Application

QM/MM methods can be performed in several ways [84],[144]: In principle it is possible to add QM attributes to an existing MM program, using the corresponding force fields, such as CHARMM [150],[151], AMBER [152], etc. They are able to simulate large amounts of atoms. Also, MM attributes can be added to QM programs such as GAUSSIAN [153], TURBOMOLE [154], etc. The named codes are able to compute electronic structures of the investigated systems and other wavefunction- or electronic density-based observables. But both strategies take only the advantages

of the parent platforms and lack flexibility in adjustment of different parameters for both MM and QM methods. Therefore, a third option, which has been used in this work, is the program CHEMSHELL [155]. It equally combines MM force fields with QM routines on an interface that is optimized for QM/MM coupling. In this work a combination of the CHARMM force field [85] together with TURBOMOLE [154] has been used for QM/MM calculations. For further computational details see section 5.1.3.

QM/MM calculations can be used to gain the energy of the system of interest and, furthermore, to obtain several properties as discussed in the sections concerning MM and QM, respectively. In this work QM/MM methods are used for geometry optimizations of large proteins including cofactors and subsequent frequency and intensity calculations for the generation of Raman-spectra. The theoretical background on vibrational spectroscopy in general and Raman-spectroscopy in detail is discussed in the next section and also in the sections 5.1.6. and 5.3.

4.4. Raman spectroscopy and computation

In this section the general principles of Raman scattering (resonant and non-resonant), as well as computational details for the calculation of Raman-spectra will be discussed. The main source of literature for this section and all subsections has been [156],[157],[158].

An inelastic scattering process of photons from matter, known as Raman effect [159] or Raman scattering, is initiated by exposure of the investigated system to electromagnetic radiation. In conventional absorption-processes, electrons in the material are excited *via* incoming photons to a higher electronic level. In contrast to this, in Raman processes the photons are not absorbed by the material, but are instead scattered from the molecule [159],[156],[157].

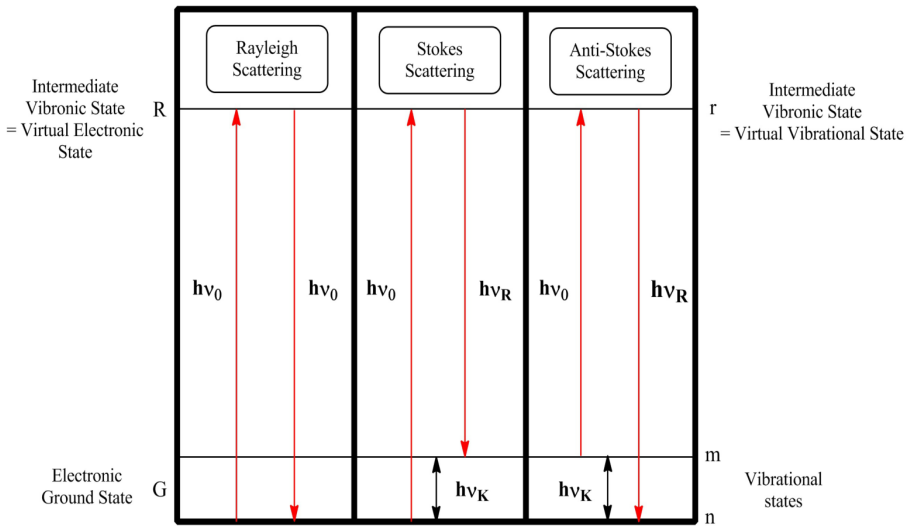

Figure 18. Schematic representation of Raman scattering processes, with corresponding energies of incomming light ($h\nu_0$), scattered light ($h\nu_R$) and observed vibration ($h\nu_K = h\nu_0 +/- h\nu_R$); G is the electronic ground-state; n, m the initial and final vibrational states and R, r are virtual electronic and vibrational states, respectively which resemble an intermediate vibronic state.

Scattering processes

In Figure 18 the different scattering processes [156],[157] are depicted. Most of the light is dispersed in an elastic way, called Rayleigh-scattering. No shift in frequency between incoming and

outgoing light can be observed and thus no energy is transferred from or onto the molecule. But a small percentage of the incoming photons are causing so-called Stokes- or Anti-Stokes-scattering processes. These are inelastic in nature. Resulting from a vibrational transition in the studied molecule, a frequency-shift between incoming and outgoing light is induced. This is named either Anti-Stokes (rise in outgoing light frequency) or Stokes (lowering of outgoing light frequency)**.**

Illustrated in Figure 18, $h\nu_0$ denotes the energy of the initial photons. In case of Stokes-scattering, the system is excited from an initial electronic (G) and vibrational (n) ground-state to a virtual vibronic state (R,r). Subsequently it falls back down to the electronic ground-state, but remains in a higher (1st excited) vibrational level (G,m). The scattered light has the energy $h\nu_R=h\nu_0-h\nu_k$, whereas $h\nu_k$ denotes the difference between the vibrational ground-state (n) energy and the one of the induced vibration k (respective state: m).

Due to higher temperatures, some molecules show already a population of the first vibrational level (G,m), even before the exposure to light. Irradiation with light then can lead to Anti-Stokes-scattering, in which again the incoming light induces an excitation from the first vibrational level (G,m) to a virtual state (R,r) in the molecule. From this it falls back to the electronic and vibrational ground-state (G,n). In this case a rise in the energy of the scattered light $h\nu_R=h\nu_0+h\nu_k$ is observed.

The frequency shifts are identical in both Stokes- and Anti-Stokes-scattering, since they always attribute to the same vibrational levels. The intensities however differ between both processes, depending on the temperature and thus on the initial state-population. Since at room temperature the population of the ground-state is more probable, Stokes processes are also more common and thus the observed intensities are often higher than for Anti-Stokes under these conditions. The computed Raman-spectra in this work are always referring to Stokes-scattering processes.

The Raman-spectrum is constituted by the scattered photons which are measured with appropriate spectroscopic techniques (not discussed here) [156],[157]. The intensity of the measured radiation ($h\nu_R$) is equal to the probability of the transition. The frequency ν_R of the measured light defines the energy of the vibrational transition and is given in reciprocal cm^{-1} and is called Raman shift. Sections 4.4.2. and 4.4.3. show the computational treatment of both frequencies and intensities of Raman scatterings [158]. Finally the bandwidth in the Raman-spectrum [156],[157] is connected to the lifetime of a the transition and is further discussed in section 4.4.4.

Raman intensities and polarizabilities

A dipole moment μ_{ind} is induced in the molecule, upon irradiation of the matter with an electromagnetic field E_0 as depcited in (66). μ_{ind} is then proportional to the polarizability α of the molecule and dependent on the frequency of the incoming light and E_0 [156],[157].

$$\vec{\mu}_{ind}=\alpha(\nu)\cdot\vec{E}_0\cos(2\pi\nu_0 t) \qquad \textbf{(66)}$$

Here $\alpha_{\rho\sigma}$ can be described as a tensor, in which the derivatives of the polarizability over the normal coordinate space for each polarization-plane(-direction) are gathered (67) [156],[157],[158]. See section 4.4.3. for more information on the polarizability tensor and its quantum-chemical treatment.

$$\alpha_{\rho\sigma}=\begin{pmatrix}\alpha_{xx} & \alpha_{xy} & \alpha_{xz}\\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz}\\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz}\end{pmatrix} \qquad \textbf{(67)}$$

The polarizability tensor $\alpha_{\rho\sigma}$ is an important observable for the description of Raman scattering processes. It is proportional to the Raman cross section σ which again is proportional to the Raman intensity I (eq. (68)) [156],[157],[158].

$$\sigma_{n\to m} \propto (\nu_0 \pm \nu_k)^4 \cdot \sum |\alpha_{\rho\sigma}|^2 \propto I_{n\to m} \tag{68}$$

Here n and m denote the initial and final vibrational level, respectively.

The polarizability tensor $\alpha_{\rho\sigma}$ then can be written according to Kramers and Heisenberg [156],[157], [158],[160] as follows:

$$[\alpha_{nm}]_{\rho\sigma} = \frac{1}{h} \sum_{R,r} \left(\frac{\langle nG|M_\rho|Rr\rangle \langle rR|M_\sigma|Gm\rangle}{\nu_{Rr} - \nu_k - \nu_0 + i\,\Gamma_R} + \frac{\langle rR|M_\rho|Gm\rangle \langle nG|M_\sigma|Rr\rangle}{\nu_{Rr} - \nu_k + \nu_0 + i\,\Gamma_R} \right) \tag{69}$$

The initial and final intermediate vibronic (electronic and vibrational) states are represented by <| and |>, respectively. Here, capital letters imply electronic states, while small letters give vibrational states. Furthermore, n, m are the initial and final vibrational levels of the electronic ground-state G. R,r are the respective electronic and vibrational intermediate states. In the case of non-resonant Raman scattering, these are virtual levels. This is in contrast to resonance Raman processes, in which the intermediate state is corresponding to the first excited electronic level (s. section 4.4.1. and Figure 19). Therefore, <||> describes a vibronic transition. The different fractions of the dipole moment operator are given by M, with and directions σ and ρ. The lifetime of the virtual state is accounted for by the damping factor Γ. ν_{Rr} is the frequency of the transition from initial to virtual states, while ν_k displays the difference between vibrational ground-state and first level. ν_0 accounts for the incident light.

Thus, the polarizability tensor can be described as a sum over all possible states and transitions. With the first summation term representing resonant contributions (s. section 4.4.1.), while the second term attributes for non-resonant transitions.

4.4.1. Resonance Raman spectroscopy

In resonance Raman processes [156],[157], the incident photons and the initiated excitations are in resonance with an electronic transition. The general principle is shown in Figure 19 for resonance Stokes-scattering, compared to the conventional non-resonant one. In resonant Raman scattering processes, the electrons are excited to an intermediate vibronic state, representing the first excited electronic state R (with r_1, r_2 the corresponding vibrational levels). The virtual levels of the transition in non-resonant Raman scattering are denoted I (virtual electronic level) and i (virtual vibrational state), respectively.

Resulting from the approach of Heisenberg and Kramers [156],[157],[158],[160], equation (69) can be simplified by omitting all non-resonant contributions, yielding equation (70):

$$[\alpha_{nm}]_{\rho\sigma} \simeq \frac{1}{h} \sum \left(\frac{\langle nG|M_\rho|Rr\rangle \langle rR|M_\sigma|Gm\rangle}{\nu_{Rr} - \nu_k - \nu_0 + i\,\Gamma_R} \right) \tag{70}$$

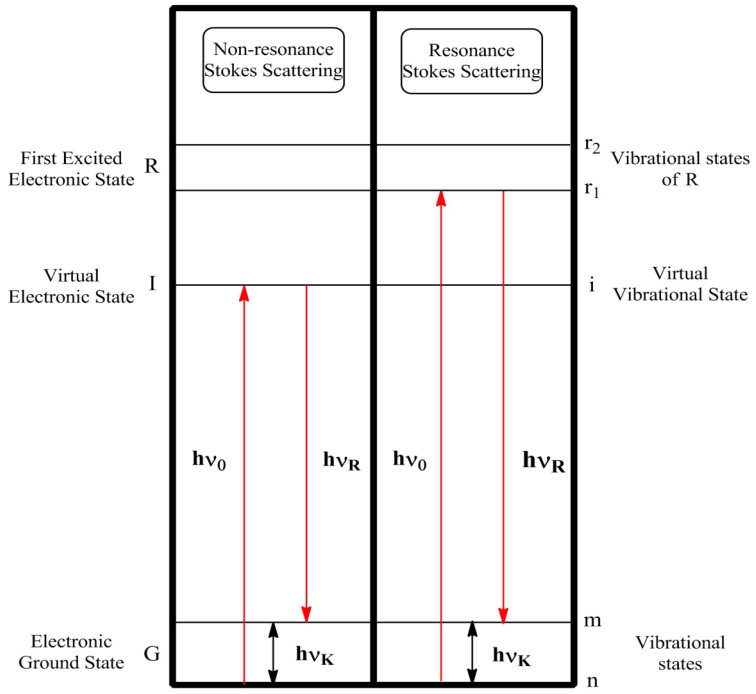

Figure 19. Comparison between non-resonance Stokes-scattering processes and resonance Stokes-scattering; Resonance Raman occurs, when the excitation is in resonance with a higher electronic level (R,r) with corresponding energies of incoming light ($h\nu_0$), scattered light ($h\nu_R$) and observed vibration ($h\nu_K = h\nu_0 - h\nu_R$); G is the electronic ground-state; n, m the initial and final vibrational states and I, i virtual electronic and vibrational states which resemble an intermediate vibronic state for non-resonant processes; R, r is the intermediate vibronic state for resonance Raman scattering with R the first excited electronic state and r_1 and r_2 the respective vibrational levels of R.

Furthermore, electron-excitations are much faster than the corresponding nuclear motions (electrons move much faster than nuclei). This means that the vibrational states are not changing during the electronic excitation process. Thus the Born-Oppenheimer approximation can be applied, resulting in a separation of electronic and vibrational transitions [107],[156],[157],[158] (71):

$$\langle nG|M_\rho|Rr\rangle = \langle nr\rangle\langle G|M_\rho|R\rangle = \langle nr\rangle M_{GR,\rho} \qquad (71)$$

Here <nr> is the so-called Franck-Condon factor, describing the overlap between both initial and intermediate vibrational states. A transition will be most probable, if the vibrational states of both electronic ground-state and excited state are similar to each other, following the Franck-Condon principle [156],[157].

The second term $\langle G|M_\rho|R\rangle$ corresponds to the pure electronic transition. It contributes to the electronic transition dipole moment $M_{GR,\rho}$ which is further expanded in a Taylor series. Combination of eq. (70) with eq. (71) then yields the expression (72) for the polarizability tensor of resonance Raman processes [156],[157],[158]:

$$[\alpha_{nm}]_{\rho\sigma} \simeq A_{\rho\sigma} + B_{\rho\sigma} \qquad (72)$$

$[\alpha_{nm}]_{\rho\sigma}$ then consists of a number of individual terms [156],[157],[158]:

A-term scattering: Depends on the electronic transition dipole moment of the Taylor series of $M_{GR,\rho}$. It becomes important for strong resonant electronic transitions, in which the excited states are largely displaced from the ground-state.

B-term scattering: Depends on the second element of the Talyor expansion of $M_{GR,\rho}$ which is the first derivative of the electronic transition dipole moment over the normal coordinate space. It is only important for transitions with strong vibronic coupling, where the conical intersection is met.

In the following sections it will be discussed how to accomplish computations of Raman-spectra [158]. In detail it will be discussed how frequency calculations (section 4.4.2.) and generation of Raman intensities (section 4.4.3.) and bandwidth' (section 4.4.4.) will be carried out.

4.4.2. Frequency calculations *via* normal mode analysis (NMA)

Fundamental for frequency computations is the normal mode analysis (NMA) [63] as used in this study. There are also other methods (such as Fourier Transformations), but this work will focus solely on the NMA approach, in which the vibrational Schrödinger equation will be solved [63], [158]:

The geometry of the investigated molecule has to be near to the energetic equilibrium structure. Therefore, the first derivative of the potential V over all spatial coordinates x_i has to be zero as shown in (73). This can be achieved by various optimization procedures of QM programs [158] (not explained here), yielding the minimized structure in cartesian coordinates and has to be done preliminary.

$$\left(\frac{\partial V}{\partial x_i}\right)=0 \qquad \textit{(73)}$$

Furthermore, the corresponding potential energy V can be approximated by a Taylor series for small deflections from the energetic minimum. It is truncated after the second quadratic term, yielding the harmonic approximation as displayed in equation (74) which can then be used to compute the force constant matrix (78) [63],[158]:

$$V=V_0+\sum\left(\frac{\partial V}{\partial x_i}\right)_0 x_i+\frac{1}{2}\sum\left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_0 x_i x_j \qquad \textit{(74)}$$

The NMA procedure [63],[158] requires internal coordinates. Therefore, common QM-programs, like GAUSSIAN [153], convert the optimized structure from cartesian coordinates into a set of internal coordinates (bonds, angles, torsions and impropers) which are used for frequency calculations according to eq. (76). These are more intuitive for description of molecular motions. Cartesian coordinates x are related to internal coordinates s *via* the Wilson B matrix [161]:

$$B=\left(\frac{\partial s}{\partial x}\right) \qquad \textit{(75)}$$

Internal coordinates can also be combined into natural internal coordinates, as used by the program GVA [162], by the following formula:

$$r = \sum_k c_k s_k \qquad \textbf{(76)}$$

According to Pulay et al. [163] all internal coordinates s, which are necessary for description of a specific normal mode, are combined.

Furthermore, following the Wilson Decius formalism [63],[156],[157],[158],[161] one gets:

$$|F - G^{-1}\lambda| = 0 \qquad \textbf{(77)}$$

Equation (77) can be transformed into the conventional normal mode equations as shown in eq. (78):

$$(GF)L = \lambda L \qquad \textbf{(78)}$$

G is the so-called Geometry matrix (eq. (79)) describing a combination of all internal coordinates s_{ta} (with atom α and internal coordinate t, t', respectively) of the molecule, with the reciprocal mass μ.

$$G_{tt'} = \sum_{\alpha=1}^{N} \mu_\alpha \cdot \vec{s}_{t\alpha} \cdot \vec{s}_{t'\alpha} \qquad \textbf{(79)}$$

The force constant matrix F is the second derivative of the potential V of the molecule over internal coordinates (eq. (80)).

$$F = \left(\frac{\partial^2 V}{\partial s_i \partial s_j} \right)_0 \qquad \textbf{(80)}$$

Together with Pulay's approach [158],[163], the force-constants can additionally be scaled, according to (81):

$$(F_{ij})^\sigma = \sqrt{\sigma_i}(F_{ij})\sqrt{\sigma_j} \qquad \textbf{(81)}$$

This will be done for the usage of the program GVA [162], as described in the methods in section 5. In equation (81) σ denotes the scaling factors and F_{ij} the respective force constants. The scaling factors are produced for a group of internal coordinates and a certain QM method and are transferable between similar molecules.

The eigenvalues λ in (78) describe the frequencies ν of the corresponding vibrations/transitions as shown in equation (82) with c being the constant of speed of light and λ the wavelength:

$$\lambda : \nu = \frac{1}{2\pi c}\sqrt{\lambda} \qquad \textbf{(82)}$$

Furthermore, the eigenvector of this equation is the so-called L-Matrix. It is important for the transformation from internal coordinates S to normal coordinates Q (83) and the generation of the potential energy distribution (PED).

$$L : Q = L^{-1} S \qquad \textbf{(83)}$$

In this work the NMA method [63],[158] has been carried out by using GAUSSIAN [153] and application of the program GVA [162] in a QM and QM/MM framework (s. section 5).

4.4.3. Raman intensities

According to equation (68) the corresponding intensities can be computed using (84) [156],[157], [158]:

$$I_k^{Ra} \propto \sigma_k \propto \frac{(\nu_0 - \nu_k)^4}{\left(1 - \exp\left[\frac{-hc\,\nu_k}{K_B T}\right]\right)} \sum |\alpha_{\rho\sigma}|^2 \qquad \textbf{(84)}$$

The Raman cross section σ_k is proportional to the frequency of the vibrational mode ν_k and the polarizability tensor $\alpha_{\rho\sigma}$ of the investigated system. Therefore, knowing the frequencies of the corresponding normal modes (solutions of NMA) and the polarizability derivatives will yield the Raman intensities I^{Ra}_k.

The computation of intensities can then be carried out as follows [158]:

In general, the cartesian polarizability derivatives are calculated (with ρ and σ being the polarization directions and H the Force constant matrix in cartesian coordinates and E the energy) according to (85):

$$\alpha_{\rho\sigma} = \left|\frac{\partial \alpha}{\partial x}\right|^2 \propto \left|\frac{\partial^2 H_x}{\partial E_i \partial E_j}\right|^2 \qquad \textbf{(85)}$$

Afterward, cartesian polarizabilities are transformed into normal coordinate Q_k-dependent ones *via* B- and L-matrices.

$$\left(\frac{\partial \alpha}{\partial Q_k}\right) = \sum_{i=1}^{3N-6} L_{ik} \sum_{j=1}^{3N} B_{ij} \left(\frac{\partial \alpha}{\partial X_j}\right) \qquad \textbf{(86)}$$

The Raman activities are computed according to equation (87):

$$A_k^{Ra} = \left[45\alpha_k^2 + 7\gamma_k^2 \right]$$
(87)

α_k and γ_k are isotropic and anisotropic parts of the polarizability tensor (polarizability derivatives).

The intensity is then computed according to equation (88):

$$I_k^{Ra} = I_0 \frac{(\nu_0 - \nu_k)^4}{\left(1 - \exp\left[\frac{-hc\,\nu_k}{K_B T}\right]\right)} A_k^{Ra}$$
(88)

The last step is only carried out, when using GVA [162] for computation. GAUSSIAN [153] only uses Raman-activities [158].

4.4.4. Bandwidth

The bandwidth $\Delta\nu$, which is proportional to the energy of the transition ΔE, is connected to the lifetime Δt of the excited vibrational states (i.e. the corresponding transitions) via the Heisenberg uncertainty principle [164] as depicted in eq. (89) [156],[157]:

$$\Delta E \, \Delta t \geqslant \frac{h}{4\pi}$$
(89)

According to (89) a short lifetime Δt (e.g. Δt=1 ps) leads to a big ΔE and thus a broad band ($\Delta\nu$=2.7 cm^{-1}) and *vice versa* (Δt=1 ms leads to $\Delta\nu$=2.7 x 10^{-9} cm^{-1}).

Since the lifetime of a vibrational transition in a molecule is strongly influenced by the environment it is in, the shape and width of certain band gives information about structures and dynamics in the investigated system. Therefore, in the current work two different approaches are used for obtaining the bandwidth:

1. The bandshape is modeled by a Lorentzian function (with fixed bandwidth $\Delta\nu$=12 cm^{-1}), connected to the corresponding intensities/activities (this approach has been used for Raman-spectra computation for *Rc*FDH cofactor) [156],[157],[158].

$$I(\nu) = \frac{I_{max}}{1 + \left(\frac{\nu_0 - \nu}{\frac{\Delta\nu}{2}}\right)^2}$$
(90)

2. A second possibility to gain access to dynamics and thus bandwidth in Raman-spectra is to generate so-called sum-spectra [158]: In these different snapshots of a MD-simulation are

geometrically optimized *via* QM/MM techniques and of each structure a Raman-spectrum is generated as described above. Afterward, the individual spectra are summed into a single one. Thus the influence of different conformations of the system at different timesteps is taken into the computed Raman-spectrum.

Chapter 5

System preparation and protocols for calculations

5.1. Computation protocols for Cph1

5.1.1. Model building for Cph1

Protein structure models based on X-ray measurements [10] were used as a starting point for the model building. The initial coordinates for the protein were taken from the Protein Data Bank [15] (code 2VEA; resolution, 2.45 Å). The model for Cph1, used in this work, was obtained from Grazia Daminelli [64]. It was already completely prepared, according to the procedure described in earlier work [165],[166]. For more details on model building, please see section 5.2.1.

Preliminary work by Grazia Daminelli has been [64]: In general, missing amino acids have been inserted manually and/or by secondary structure prediction algorithms [167]. Absent hydrogen atoms have been added to the heavy atoms via the HBUILD [168] routine, implemented in CHARMM [150],[151]. Protonation states for histidines have been evaluated via visual inspection of the surrounding environment. Exception have been the two histidine residues in near vicinity of the PCB cofactor, namely His260 and His290. According to the results of previous NMR studies [29] these two have both been protonated only on the ε-nitrogen. Charged amino acids (lysine, arginine, aspartate and glutamate) were protonated according to standard CHARMM [150],[151] protocol. A hexagonal water box consisting of TIP3 water [101] with the dimensions of a=b=119.04 Å, c=146.2 Å; α=β=90.0°, γ=120.0° has been built around the protein. Chloride or sodium ions have been placed within 3 Å of each amino residue not involved in a salt-bridge. The system has been neutralized by random placing of ions [166]. This was necessary to obtain a system charge of zero which is required for Particle-Mesh-Ewald calculations [96],[97]. The protonated, solvated and neutralized system has then been used for MD-simulations, as described in the next subsection.

5.1.2. MD-simulation protocol for Cph1 with classical force field

The NAMD [169]-package (version 2.6) together with the CHARMM22 force field [85] has been used for MD-simulations according to the protocol displayed in Figure 20a. The Particle-Mesh-Ewald method [96],[97] has been applied to compute electrostatic interactions, while van der Waals forces have been treated with switching functions (switching distance of 10 Å) and a cutoff of about 12 Å. Periodic boundary conditions (PBC) have been used for the hexagonal cell. Further a time step of 2 fs has been applied, together with the constrain algorithm SHAKE [170]. It has been used to guarantee minimum values for all bond lengths between heavy atoms and hydrogen. For the chromophore PCB, previously optimized parameters [171] have been used. In the productions run all heavy atoms were completely free of any constrains. A Noose-Hoover Thermostat [171],[172] was used to maintain NPT conditions in all productions runs.

The MD-simulations have then been carried out as follows:

1.) Preliminary work like model building.

2.) Solvent water molecules (as well as hydrogen atoms) are minimized (4 ps) to remove bad contacts, then heated for 30 ps to a temperature of 300 K under stepwise release of restraints of the

water molecules. Afterward water/hydrogen atoms are equilibrated for 50 ps with no restrains for TIP3-water and hydrogen. All for several thousand steps. During step 2, all other heavy atoms have been kept fixed. Step 2 has been preliminary carried out by Daminelli et al. [64].

3.) The resulting structure from 2.) is then used in an all-atom energy minimization (20 ps) under constraints for all heavy backbone atoms of the protein and the chromophore with a harmonic constrain of 15 kcal/mol $Å^2$. These are removed stepwise, until all atoms are completely free.

4.) The minimized system is then heated to 300 K for 200 ps. Again heavy atoms of protein backbone, as well as chromophore are restraint. The atoms are gradually freed, until they are completely unrestraint.

5.) The heated and completely unrestrained system is then further equilibrated for 200 ps with constant pressure and temperature. To obtain NPT conditions the Noose-Hoover thermostat together with the Langevin method [171],[172] as provided by the program NAMD [169] can be used.

6.) Finally a production run of 40 ns (consisting of 200 0.2 ns steps) of the completely unrestraint system is done.

The resulting molecular structures and observables from the MD-trajectory, like RMSD and RMSF can be visualized and computed via the Visual Molecular Dynamics software VMD 1.8.6. [173].

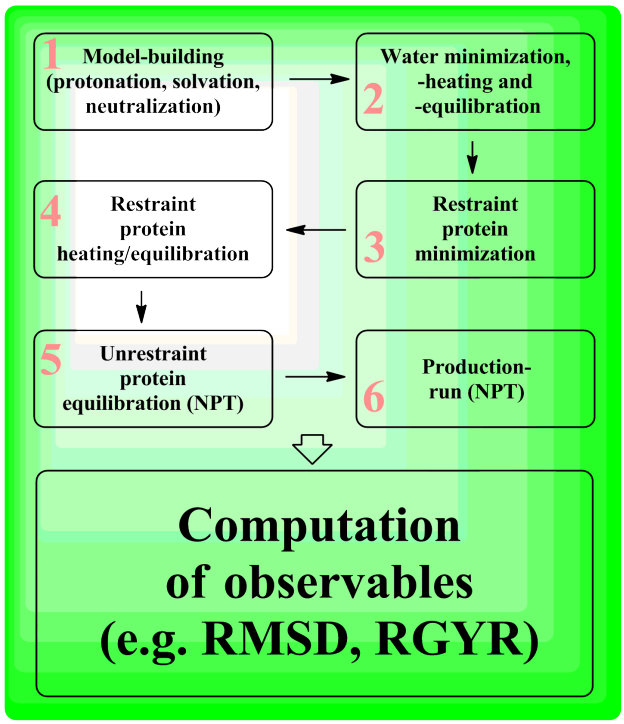

Figure 20a. General MD-simulation receipt

5.1.3. Generation of a polarized force field for Cph1 and computation of Mulliken charges

For the generation of a polarized force field for Cph1 a QM/MM method has been used. In detail, a SCC-DFTB [117]/CHARMM force field [85] approach was used to compute the partial charges of all residues which have at least one atom inside a sphere with a radius of 4 Å from the center of

PCB (s. Figure 20b, right). The remaining part of the system was treated by classical molecular mechanics. Additionally, the partition of the system in a QM and a MM fragment was defined by specifying those C-C single bonds, at which the covalent linkage between two fragments is cut using the SPLAM [174] method. For all selected residues the C_α-C_β bond was the cut one. Therefore, only the respective side chains have been treated by SCC-DFTB. The computations have been carried out using the CHARMM32 program [150],[151] and optimized parameter sets (Hamiltonian approximations done with DFT in a minimal basis and with GGA functional) developed by the group of Elstner et al. [175] which are optimized for implementation of SCC-DFTB calculations into CHARMM. Furthermore, the MM-electrostatic potential has been imported into the QM-Hamiltonian [174]. Finally, the Mulliken [98] charges have been calculated for each enlisted residue *via* an iterative procedure until charge convergence was reached. These were then exchanged with the corresponding old charges in the psf-file to generate a polarized force field.

PROTEIN-RESIDUE	COFACTOR	WATER
LEU 15	PCB	HOH 2
LEU 18		HOH 3
ILE 20		HOH 4
MET 174		HOH 10
TYR 176		HOH 11
VAL 186		HOH 28
TYR 198		WAT 632
TYR 203		WAT 663
SER 206		WAT 566
ASP 207		WAT 572
ILE 208		WAT 577
PRO 209		WAT 578
ALA 212		WAT 579
LEU 215		WAT 586
PHE 216		WAT 589
ARG 222		WAT 605
ILE 224		WAT 614
ARG 254		WAT 610
ALA 256		WAT 619
TYR 257		WAT 620
HSE 260		WAT 624
TYR 263		WAT 626
MET 267		WAT 629
SER 272		
THR 274		
LEU 286		
ALA 288		
HSE 290		
ALA 457		
TYR 458		
LEU 469		
HSD 470		
PRO 471		

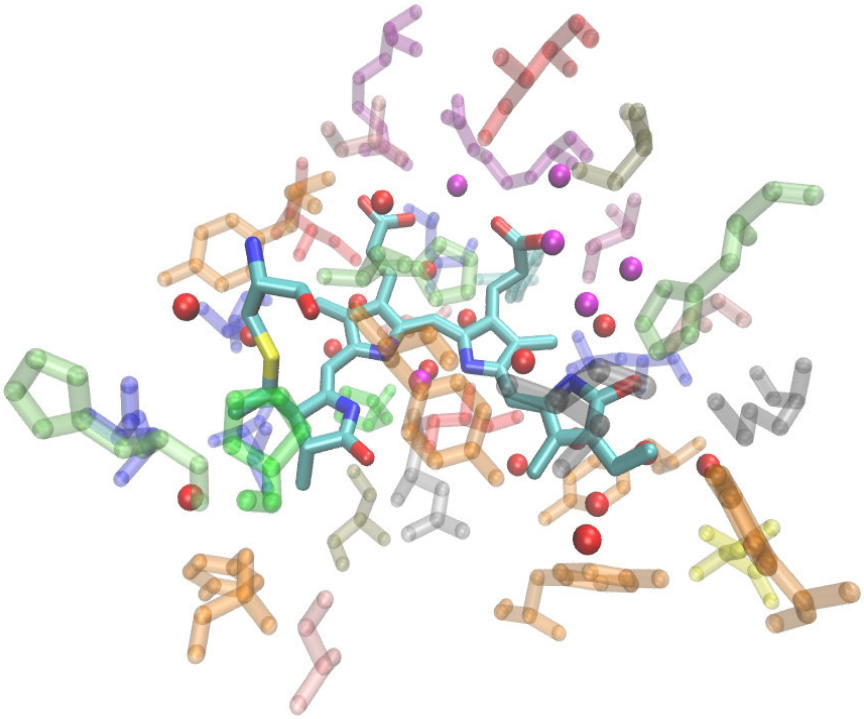

Figure 20b. Residues involved in QM-Mulliken-charge computation (hydrogens not shown; protein residues in ghostly representation); HOH means crystal-water and WAT denotes solvent water

5.1.4. MD-simulation protocol for Cph1 with polarized force field

MD simulations with polarized force field (polff) have been carried out in an identical way as for MD-simulations for Cph1 with classical force field (see section 5.1.2.; steps 1 to 6), using a psf structure file containing the polarized charges of the selected residues.

5.1.5. QM/MM geometry optimizations for Cph1

QM/MM model building

As starting points for the QM/MM calculations 25 snapshots (of every 4 ps) of the last 100 ps of both the MM- and polff-MD trajectory have been taken. The water-box has been cut down to a sphere with a radius of 40 Å centered on ring C of the chromophore. The QM/MM optimization procedures were according to recent publication [63] and as discussed in section 4.3. The system has been divided into a QM and a MM region, with a boundary region in between, according to the additive scheme [84],[143]. The boundary region was accounting for coupling between QM- and MM-atoms. The general concept is visualized in Figure 16 in section 4.3.

The QM region consisted of the protonated bilin cofactor (PCB), the pyrrole-water and the side-chain of protein-attachment Cys259 at the cofactor. The corresponding atoms were treated at a quantum-chemical level. All other residues/atoms were either treated by MM force field techniques or remained fixed during the optimization.

QM/MM procedure

The CHEMSHELL program [155],[176] has been used for all QM/MM calculations, combining the TURBOMOLE package [154] for the QM-related parts and CHARMM [85] for the MM-part. The actual calculations have been carried out according to the following scheme:

1. QM atoms are computed using conventional DFT methods [108],[110] with the B3LYP functional. A 6-31G* basis set has been applied to all heavy atoms, while hydrogen atoms have been treated with a 4-31G* basis set. DFT-calculations were carried out by TURBOMOLE [154].

2. The CHARMM27 force field [85] is used for description of the MM part of the system. The latter is further divided into an active and an outer region.

3. All atoms inside a 15 Å sphere centered at the pyrrole-nitrogen of ring C of the chromophore PCB are considered active during the QM/MM optimizations and contribute to the active region. All atoms outside the active region remain fixed during the whole optimization. They are treated as point charges. The region outside the 40 Å water sphere is called outer region and contains no explicit solvent molecules. Additional forces to keep the explicit water molecules inside the active region are generated by the point charges of the atoms outside the active region.

4. Bonds crossing the border between QM and MM region are treated by the link-atom scheme (as discussed in section 4.3.3.). Respective QM-atoms of the cut bonds are saturated with hydrogen-atoms, while for the MM-atoms a charge-shift scheme is applied [84],[143],[144].

5. The electrostatic embedding method is used for the electrostatic interactions between QM and MM atoms [84],[143],[144].

6. The optimization procedure is carried out with the help of a limited memory quasi-Newton L-BFGS algortihm [63]. Furthermore, hybrid delocalized internal coordinates (HDLC) [177] are used.

The results have been visualized with VMD [173].

The QM/MM optimized snapshots have then been used for further Raman-spectra computations according to the NMA approach [63],[158], as described in the next subsection.

5.1.6. Raman-spectra calculation with NMA for Cph1

The QM/MM geometry optimized structures in cartesian coordinates have been used for subsequent frequency computations with the GAUSSIAN program [153]. The NMA approach has been used as described in the following [63],[158]:

1. The cartesian structure is converted into a set of internal coordinates according to equation (81). The structural information in internal coordinates is then used to generate the G-matrix. Natural internal coordinates [158] have to be used in case of subsequent usage of the program GVA [162]. These have to be manually prepared according to Pulay's approach [163], as depicted in equation (82).

2. The force constant matrix is computed as the second derivative of the potential energy of the system as provided by the previous geometry optimizations. The force constant matrix in cartesian coordinates (called the Hessian). H is then converted into internal coordinate force matrix F. This is carried out according to F=BH, with B the Wilson B-matrix [158],[161].

3. Afterward, the vibrational Schrödinger equation (76) is solved by diagonalization of GF. This yields the fundamental frequencies of the normal modes. The eigenvectors L (amplitude of motions) will be used for computation of the potential energy distribution (PED) according to $P=L^{t}L^{-1}$. It displays how much an internal coordinate is contributing to a normal mode [158].

4. Raman-activities were computed according to equation (87) with use of prior computation of the polarizability derivatives (eq. (85)) and conversion from cartesian to normal coordinate representation (eq. (86)) [158]. It has to be noted that these are not directly comparable to experimental Raman-intensities.

Steps 1. to 4. have been carried out with the QM-package GAUSSIAN [153].

5. Subsequently, GVA [162] has been used for obtaining normal mode frequencies and a PED in a more accurate fashion. The difference in computational protocol is that the optimized cartesian coordinates are manually transformed into a set of natural internal coordinates. Furthermore, since in QM/MM geometry optimizations an additional link-atom has been used, this is taken out of the computed Hessian. Both vibrational frequencies and Raman-intensities are computed by GVA.

The advantage of GVA-produced output is a better and “easier to interpret” PED [162]. The PED obtained by frequency calculations with GAUSSIAN [153] is actually only interpretable for small and easy molecules/systems (as was done for the cofactor-studies of *Rc*FDH, as described in section 5.3.). Furthermore, the force-constants can be scaled in an efficient way according to equation (83). This may be helpful to overcome shortcomings of the harmonic approximation, as well as of the used quantum mechanical method which may produce systematic errors in the computed vibrational spectra.

The final Raman-spectra are plotted, using the computed frequencies and intensities with the help of a Lorentzian function with a bandwidth of 12 cm^{-1}. It should be further noted that in accordance to previous studies [178],[179], the comparison of intensities between GVA [162]-computed Raman-spectra and experimental Resonance Raman spectra of tetrapyrrolic chromophores is valid.

5.2. Computation protocols for Agp2

5.2.1. Model building for Agp2 wild type (WT)

An early unpublished version of the X-ray structure from Agp2 Pfr state (with an endo-cyclic double bond of ring A of BV) from Scheerer et al. [16] has been used as an initial set of coordinates for the computations. Missing protein residues 80-85 were identified *via* primary structure analysis. They were part of a loop region. The positions of the corresponding α- and β-carbons were chosen by visual inspection of the environment and then placed manually. Afterward, the missing atoms were included by using the CHARMM [150],[151] software, followed by a short minimization (100 steps) for reducing bad contacts. Biliverdine (BV) was bound to a native Cys *via* endo-cyclic displacement of the double-bond of ring A [16].

Since, for folding purposes, the crystals were obtained after methylation [180] of all lysine (Lys) residues, three Lys in a radius of 20 Å around BV also occurred as methylated species in the X-ray structure [16]. These three methylated residues were manually replaced by native Lys.

Protons for titratable groups have been assigned *via* visual inspection of the surrounding environment, potential hydrogen-bonding partners and saltbridges, with the exception of His248 and His278 (see last paragraph of this subsection). All missing hydrogen atoms were then inserted *via* the HBUILD routine [168] of the CHARMM code [150],[151]. Additionally, a proton has been added manually at the propionic side-chain of ring C according to experimental findings [23].

Subsequently the protein has been solvated in a hexagonal box with a=b=90, c=120 and α=β=90.0°, γ=120.0° of TIP3P water molecules with the help of CHARMM [150],[151]. Additionally, the solvated structure has been treated with NaCl (using the VMD package [173]). A sodium or chloride ion has been placed in proximity of each charged amino acid not involved in forming salt-bridges. Afterward, ions have been placed randomly in 5 Å proximity of the solute to achieve electro-neutrality of the system [166].

Finally the water molecules and hydrogen atoms in the structures were then equilibrated with help of the program NAMD [169] and the CHARMM all-atom force field [85]. First a minimization of 5000 steps was carried out, followed by heating for 1 ns and equilibrating for 15 ns (under slowly reducing restraints for water molecules). For further details of simulation procedure s. section 5.1.2.

As described in section 3.2., the protonation states of the conserved His248 and His278 are of large significance. Therefore, nine models have been built (dd, de, dp, ed, ee, ep, pd, pe, pp) in which all possible protonation states of the two histidines are combined. The former letter refers to His248 and the last one to His278. Here d denotes the protonation on N_δ and e the protonation of N_ε. P represents a doubly protonated histidine. The different protonation states were realized with the HBUILD routine [168] of CHARMM [150],[151]. All further discussed QM/MM computations were carried out for all nine models. The same has been done for six models (de, dp, ee, ep, pe, pp) with a chromophore with missing hydrogen on PSC(C). See Figure 21a for examples of built models of Agp2.

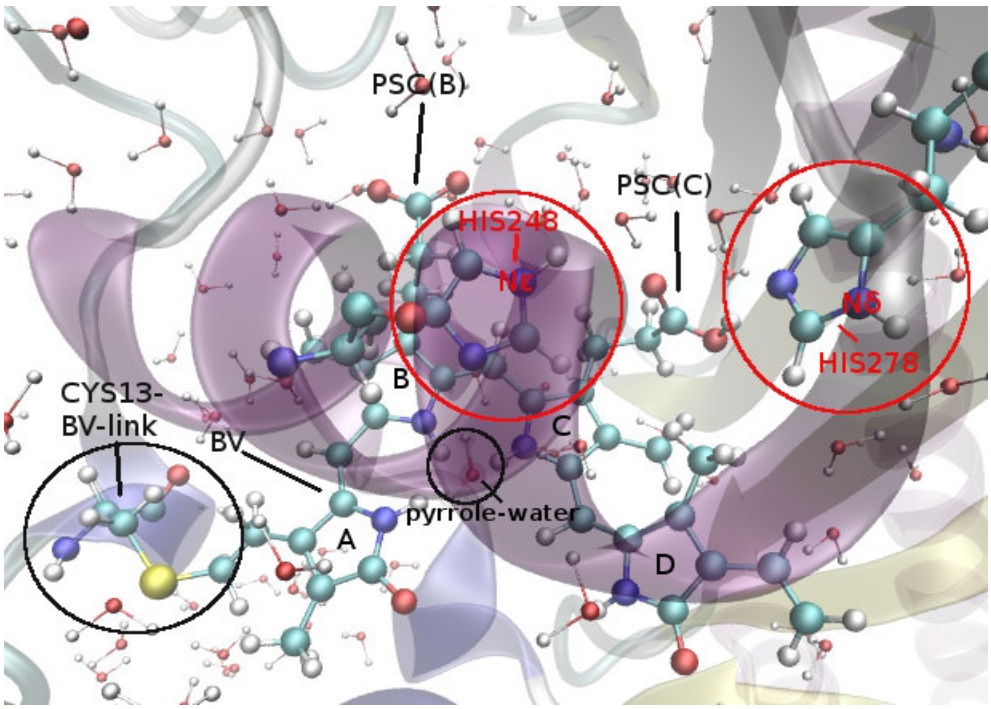

Figure 21a. Structure of Agp2 BV binding-pocket with ed-protonation of H248 and H278, solvated and neutralized (after equilibration) with surrounding solvent and crystal waters; based on [16]

5.2.2. Model building for Agp2 BV-variants

Additionally, different non-native variants of the cofactor BV have also been used inside the protein structure of Agp2. In detail, BV was either methylated at PSC(B), PSC(C) or both (see Figure 21b). The propionic side-chain which was not methylated was neutralized by adding a proton. These modifications have been included in the corresponding structures *via* patching of the protonated crystal structure [16]. Parameters for CH_3 groups have been taken from acetaldehyde [85]. The models have been protonated, solvated, neutralized and equilibrated as described in 5.2.1. The ed-model for His248 and His278, respectively, has been used for these types of BV-variants.

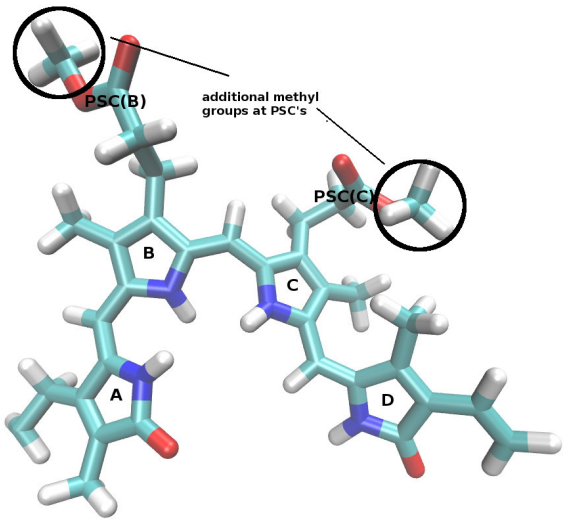

Figure 21b. Example of BV-variant: Structure of Agp2 chromophore BV with additional methyl groups at propionic chains of rings B and C

5.2.3. MD-simulation protocol for Agp2 (WT and BV-variants) with classical force field

All MD-simulations with classical partial charges for Agp2 (+ variants) have been carried out according to the protocol as discussed in section 5.1.2. Differences in the simulation conditions for Agp2 in contrast to Cph1 have been: All heavy atoms of the chromophore BV have been kept fixed during all simulation steps/procedures. This was necessary due to an inefficient parametrization. The used parameters were only fitted for the Pr state of BV and not the Pfr. Tests with free atoms of BV showed strong unnatural behavior like twisting of bond angles and generation of strained geometries between the individual rings. Water has been minimized (10 ps) , heated (0.5 ns) and equilibrated (7.5 ns) according to section 5.1.2. Afterward, minimization, heating and equilibration of the complete system has been carried out similar as in described in section 5.1.2. Finally, production runs of 20 ns have been carried out for Agp2-systems with the different variants of the chromophore.

5.2.4. QM/MM geometry optimizations for Agp2 WT

QM/MM model building

As starting point for the QM/MM calculations, the solvated X-ray structure [16] with equilibrated water-box and hydrogen atoms have been taken. The water-box has been cut to a sphere with a radius of 40 Å centered at the pyrrole-nitrogen of ring C of BV. The QM-part consisted of BV and 11 water molecules and the side-chains of protein residues His248, His278, Tyr251, Tyr205, Asp196, Ser462, Cys13 (170 atoms) (see Figure 22 for an example of the dd model).

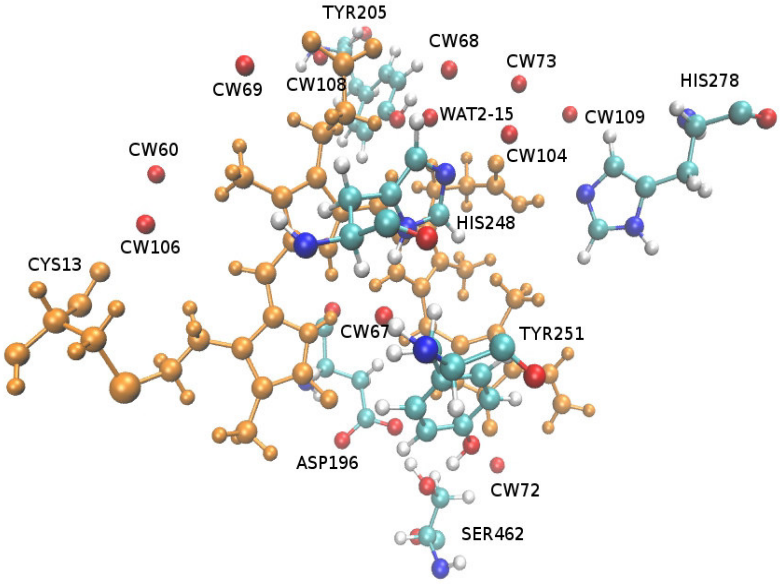

Figure 22. QM-region of dd-model of Agp2: BV+Cys-link (orange); CW (crystal-waters)+protein residues (color)

QM/MM procedure

The program CHEMSHELL [155],[176] has been used for QM/MM calculations. The QM region was tretated with DFT (B3LYP). 6-31G* and 4-31G* basis sets have been used for heavy atoms and hydrogen atoms, respectivley. The QM-package TURBOMOLE [154] has been used for the DFT-computations. MM-atoms are described by the CHARMM27 force field [85]. An active region with

radius of 20 Å centered at the N of ring C of the BV-cofactor was chosen, including the chromophore and surrounding residues and water molecules. All atoms outside the active region remained fixed during the optimization and where treated as point charges. The geometry of the chromophore binding-site for all nine models was optimized as discussed in section 5.1.5.

Additionally, following the same procedure, some calculations with extended basis set for hydrogen atoms has been carried out: Here, a 6-31+G* basis set with polarization and diffuse functions have been used for hydrogens, while the other atoms have been treated with 6-31G* as before. Due to the higher computational costs, the QM region has been reduced to the side-chains of Cys13, His248, His278, BV and the p-water and three water molecules.

5.2.5. QM/MM geometry optimizations for Agp2 BV-variants

For Agp2 BV-variants a QM-region including the side-chains of Cys13, Arg211, His278, His248, Asp196, BV-variant and the 11 waters has been chosen. The optimization procedure was the same as for Agp2 WT and as discussed in sections 5.1.5. and 5.2.4.

5.2.6. Analysis of optimized geometries for Agp2 (WT and BV-variants)

QM/MM optimized structures were analyzed using the MOLDEN [181] program for estimation of correct convergence and investigation of H-bond-networks and VMD [173] for RGYR, RMSD, RMSF and distance calculations and comparison to the protonated, solvated and water-equilibrated crystal structure [16] as reference.

5.3. Computation protocols for *Rc*FDH

5.3.1. Model building with QM software

As starting point for all used models, the homology models by Hartmann et al. [49] have been used (see Figure 23) for both the oxidized and reduced form of the cofactor. These models based on different crystal structures of FDH of *E. coli* [52],[56], Fe-only hydrogenase from *Chlostridium pasteurianum* [182] and NADH-dehydrogenase from *Thermus thermophilus* [183]. The actual model building has been done with the Gaussview [184] software.

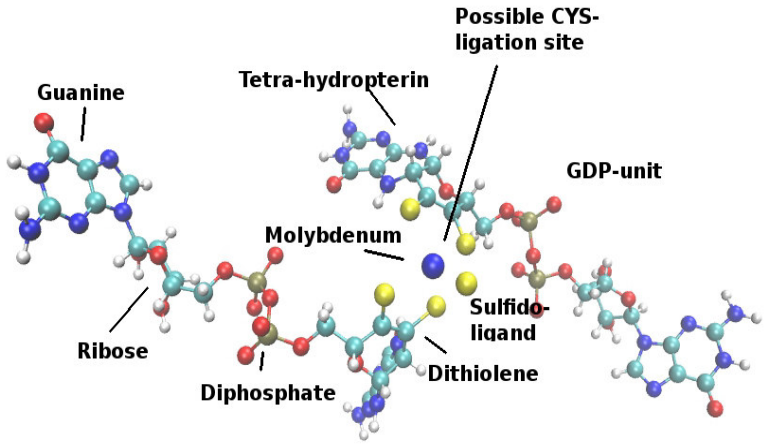

Figure 23. Mo-cofactor (homology model of oxidized *Rc*FDH [49]; Cys386-ligation is not shown)

Since the complete Mo-containing cofactro (Moco) is very large and thus computational costly, only a fragment of the Moco has been used for in most calculations (see section 12.1.4. for detailed information on model-size tests) which is named model 1. All used models are depicted in Figure 24. If not stated otherwise all models have been built in the following:

In all used model 1 compounds the GDP-units were omitted. Dithiolene ligands, as well as linked pterin moieties have been included. Models with Mo in +VI oxidation had Cys386 coordinated to it (with one exception, where Cys386 was absent for test purposes, as described in section 12.2.1.) and O, OH, S or SH ligand present. Models with Mo in +V/+IV oxidation had either Cys386 present or absent. Either Cys386 is ligated to metal or uncoordinated and either charged or protonated and in near proximity to the cofactor. In all MoIV models O, OH, S or SH is ligated as sixth ligand at molybdenum. The models are named MoZ_X, were X is either an O, OH, S or SH ligand and Z denotes the oxidation-state of Mo which can either be +VI, +V or +IV. If used, Cys386 has been mimicked by a truncated version, in which the C_β is saturated by protons. C_β remained fixed during all computations in order to simulate the lesser degree of freedom the residue would feel in the native protein environment (s. Figure 24). Further, the truncated GDP-connection-sites (s. Figure 24) have been fixed, in order to mimic the real situation in the complete Moco.

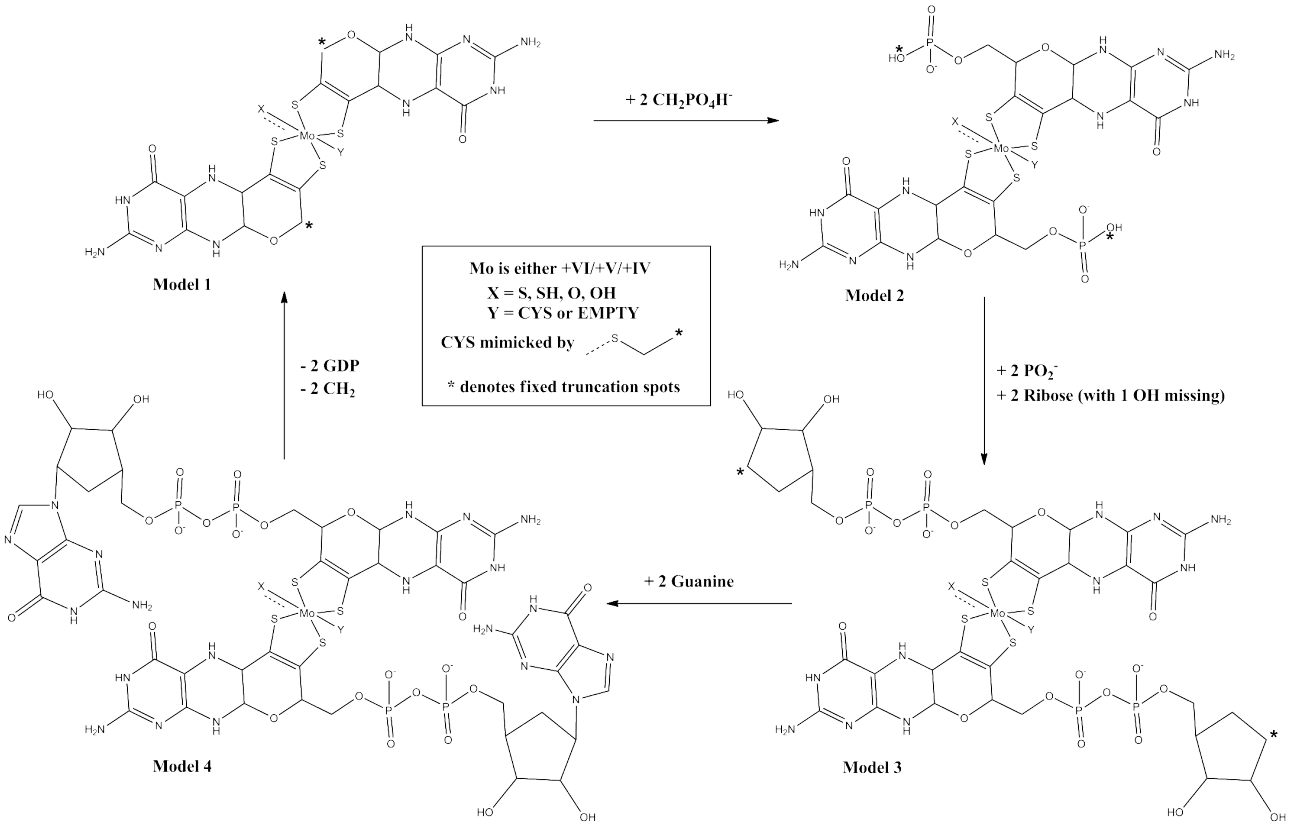

Figure 24. Schematic representation of used models 1-4 for QM; X denotes either S, SH, O or OH ligands; Y denotes either ligated Cys (mimicked by truncated version) or empty ligation site; Mo is in either +VI, +V or +IV state; asterisk denotes fixed truncation spots

Charges have been assigned as follows: each dithiolene (-2), deprotonated Cys (-1), doubly estered PO_4 (-1), O/S group (-2), OH/SH group (-1) and Mo either (+6), (+5) or (+4). Multiplicity of the complex has always been singlet for MoVI and MoIV (doubly occupied MOs) and doublet for MoV. The protein environment has always been mimicked by application of a dielectric field constant of ε=4 with PCM-calculations [136] in accordance with previous studies [71],[137],[138], [139] (computations in vacuum have also been carried out and are shown in Appendix B).

All used model 1 compounds differ only in the possible ligation of Cys to molybdenum and the identity of the sixth ligand (X in Figure 24). Models with Mo-ligations differing from the descriptions in this chapter are: Model 2 with additional CH_2+PO_4H^- attached at dithiolene and model 3 compounds with attached diphosphate and ribose at the dithiolene moieties, as well as model 4 compounds (complete Moco; only for test purposes). Furthermore, MoVI_X model 1 compounds with missing Cys386, MoV_X model 1 compounds with differing ligation-scenarios for Cys386 and MoVI-, MoV- and MoIV-models 1 with other ligands, such as N_3^- and formate. These are described accordingly in the respective sections in part V of this work and the Appendix.

The main focus in this work has been on calculations with model 1 compounds. For tests of models 2, 3 and 4 please see section 12.1.4 and the Appendix.

5.3.2. QM – calculations for *Rc*FDH models

Geometry optimizations

The geometry optimizations follow the standard procedures from GAUSSIAN09 [153], using DFT-methods [108],[110] with the BP86 [111] functional and a 6-31G* basis set for C, H, N atoms [185],[186] and def2-TVPP basis sets for S and Mo [187]. Additionally, an ECP [188] at the same level of theory has been applied to the metal atom Mo. This level of theory is in accordance to recent theoretical investigations, as done within the work of Mota et. al. [53] and Porcher et. al. [189]. Concerning the basis-set the treatment of metal and the sulfur-atoms in this work is of even higher accuracy, as compared to named works. For description of the environment of the cofactor, PCM-methods [136] have been applied, as implemented in GAUSSIAN09 [153]. Geometry optimizations have been carried out using standard SCF-convergence methods with a convergence criteria of 10^{-5}. Exceptions to this are models including azide and models with MoV. These had to be computed with a combination of reduced convergence criteria of 10^{-4} and a Newton-Raphson algorithm (quadratically convergent, QC) [158].

Raman-spectra computations

Following geometry optimizations, the vibrational frequencies and Raman-activities have been computed using the NMA-approach [63],[158] as described in section 5.1.6. (steps 1. to 4.; no application of GVA [162] and thus Pulay's approach [163]). It was carried out by the GAUSSIAN program [153]. Since no GVA procedure was used, neither the frequencies, nor the force-constants have been scaled. Also only Raman-activities have been computed, since GAUSSIAN only provided these. The obtained frequencies and activities have been plotted, using a Lorentzian function with a fixed bandwidth of 12 cm^{-1} for modeling of the bandshape [158]. The computed Raman-spectra have been compared to experimental Resonance Raman-spectra under different conditions for both oxidized and reduced species of *Rc*FDH (together with different stabilizing agents, like azide). The experiments were carried out and the corresponding data was provided by Stefan Wahlefeld et al. in the group of Prof. Dr. Peter Hildebrandt at the TU Berlin [73]. For further information on experimental procedures, as well as the respective data, please consult sections 12. and 13.

Thermodynamic binding energies

Bond-breaking energies/enthalpies have been calculated after geometrical optimization of individual models (as shown in 13.5. in detail) by the following subtraction scheme:

$$\Delta E_{bond-breaking(A-B)} = E_A + E_B - E_{AB} \quad \textbf{(91)}$$

Here, $\Delta E_{bond\text{-}breaking}$ is the energy needed to break the bond between two moieties A and B and E_{AB} the energy of the connected systems A and B, while E_A and E_B are the respective energies of the system-parts A and B, respectively. ΔE corresponds to the total electronic energies of the corresponding systems after the respective geometrical optimizations.

Furthermore, the bond dissociation enthalpies have been computed [190] using the formula (92) (with H° being the standard enthalpy; Δ_f meaning formation and Δ_r meaning reaction):

$$\Delta_f H°(298.15\text{K}) = \sum_{products} (\Delta_r H°_{product}(298.15\text{K})) - \sum_{reactants} (\Delta_r H°_{reactants}(298.15\text{K} \quad \textbf{(92)}$$

Frequency computations with GAUSSIAN09 [153] generate the sums of electronic and thermal enthalpies [190] $\varepsilon_0 + H_{corr}$ which makes it possible to compute reaction enthalpies according to (93):

$$\Delta_r H°(298.15\text{K}) = \sum (\epsilon_0 + H_{corr})_{products} - \sum (\epsilon_0 + H_{corr})_{reactants} \quad \textbf{(93)}$$

Here, ε_0 is the electronic total energy of the system and H_{corr} is the enthalpy correction. Hence, only molecular information is needed and no atomic information is necessary [190].

5.4. Computational resources

Chemical formulas have been built with the program Chemdraw [191]. Other visualizations of molecular matter has been carried out with either VMD [173], Gaussview [184] or Molden [181]. The computations have been carried out using several different resources: For MD-simulations the Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen [192] (HLRN) provided the necessary computation times. QM/MM-calculations, as well as QM-geometry optmizations and subsequent frequency computations were done with the help of the computer cluster of the Institute of Mathematics [193] at the Technische Universität Berlin. Computations of observables, such as RMSD and RMSF, as well as usage of VMD [173], Molden [181], Gaussiview [184], Chemdraw [191] and standard programs have been carried out at conventional office computers at and provided by the Technische Universität Berlin and the working group of Prof. Maria Andrea Mroginski.

Part III.

Molecular dynamics and QM/MM based Raman-spectra computations for phytochrome Cph1

Chapter 6

MD simulations of Cph1 with classical force field

MD-simulations provide a useful tool to visualize structural and dynamical properties of biological systems. It is less demanding in terms of computational costs than pure quantum mechanics and nevertheless is able to yield important insights into biological processes. It can also be combined with other methodologies, like QM/MM, to obtain for example Raman-spectra of the investigated matter. A prerequisite for this is the correct application of force fields and a realistic simulation of biological behavior. Since MD-simulations use point-charges to describe the charge environment of atoms, the influence of polarization effects is not included in conventional force fields (s. section 4.1.). This deficiency can lead to wrong or unrealistic dynamics of e.g. water molecules. As stated in section 3.1. it has been found e.g. in the case of bacteriorhodopsin that classical force fields with conventional point-charges lead to a decay of the water-network [57],[58],[59] which is in contrast to NMR-observations [60]. Especially waters near the Schiff base showed high mobility and a high rate of fluctuations (instability on its supposed position) [61],[62], where in experimental studies they have been found stable [60]. This has been attributed to the insufficient water-model which has been used [61],[62] and the missing of polarization effects in the applied force fields [65]. This has been addressed by using a so-called polarized force field which led to a stabilization of the water-network [65],[66],[67].

A similar observation has been made in studies by Mroginski et al. [63] and unpublished studies by Daminelli et. al. for *Cyanobacterial phytochrome* Cph1 [64]. They found that the highly conserved (crystal) pyrrole (p)-water showed high mobility and thus instability on its starting position between the pyrrole-rings of the PCB chromophore. The crystallographic structure [10] and NMR-studies [194],[195] suggested that this water remains on this position. This instability of named water molecule led to a broadening of the corresponding NH-in-plane (ip) rockings in calculated Raman-spectra [63] and therefore, makes them less comparable to experiments.

To overcome this problem a new set of polarized or Mulliken charges have been derived for the chromophore binding-pocket (CBD) of PCB using QM-methods in a hybrid QM/MM approach, including the chromophore and the p-water. These display the polarization effect of the environment on the cofactor and *vice versa* and have been included into the existing force field. In this way a polarized force field for Cph1 has been generated. It has been used for MD-simulations to show its superiority over classical force fields for stabilization of the p-water. Furthermore, the results of these MD-simulations have been used in subsequent hybrid QM/MM optimizations and Raman-spectra computations to show the effect of a more stable p-water on the corresponding NH-ip rocking bands of the pyrrole-rings. Thus, the necessity of polarized force fields and their development for correct reproduction of structural and dynamical properties of biological systems will be discussed. See sections 5.1.3. for details on generation of the polarized force field and 5.1.4. for the MD-simulation protocol. All used models of Cph1 were based on the X-ray structure [10] which has been taken from the Protein Data Bank [15] (entry: 2VEA, resolution 2.45 Å).

6.1. Instability of pyrrole-water with classical MM force field

RMSD of Cph1 and chromophore PCB. The highly mobile nature of the pyrrole-water with classical force fields was demonstrated by 40 ns long MM-MD simulations in NPT environment, with all free atoms, as described in section 5.1.4. First the stability of both protein and PCB were

confirmed by monitoring the average RMSD values (with respect to solvated X-ray structure [10]) over the course of the 40 ns production run. Results can be seen in Figure 25. The structure of the chromophore stays stable during the whole 40 ns run without any fluctuations in the RMSD. In contrast to this the protein shows increasing RMSD values in the last 10 ns of the simulation (up to 5 Å). They are due to movements of loop-regions, while the backbone of the protein stays stable. These fluctuations may originate in the high mobility of water-network, due to the missing description of corresponding polarization effects. The average RMSD values of the protein over the whole course for Cph1 is about 4 Å and for PCB it is 1.25 Å.

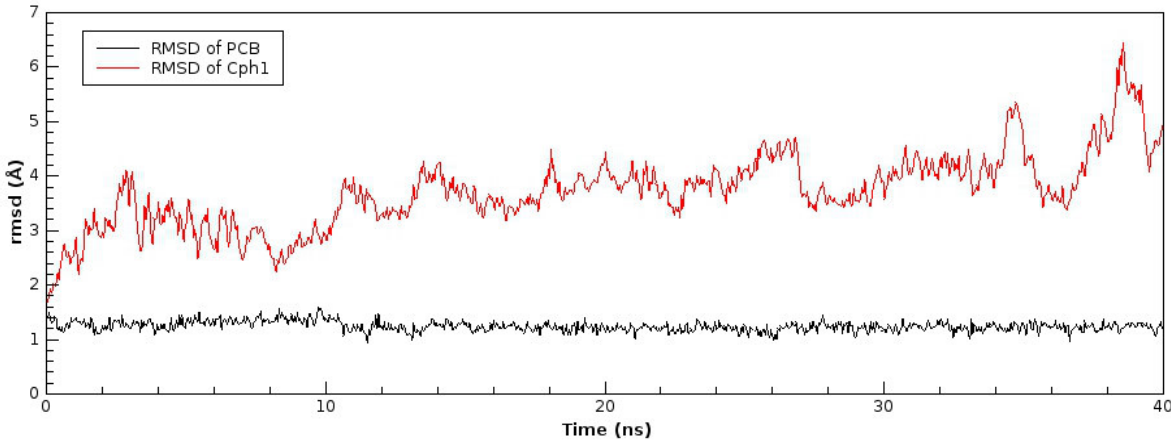

Figure 25. Average RMSD of Cph1 (red) and PCB (black) over time of 40 ns of production for classical MM-MD

Structure and RMSF calculations of PCB. The overall structure of the chromophore at the end of 40 ns (see Figure 26) shows planar rings A and C, while rings B and D of PCB are partly tilted out of plane (for starting structures and position of p-water, please s. Figure 5). RMSF calculations of the cofactor suggest, the most flexible part of PCB being ring D (RMSF=10 Å) and the methine bridge between rings C and D (RMSF=9 Å), followed by both methyl groups on rings A and C (RMSF=8 Å) and the methine bridge between rings A and B (RMSF=7.5 Å).

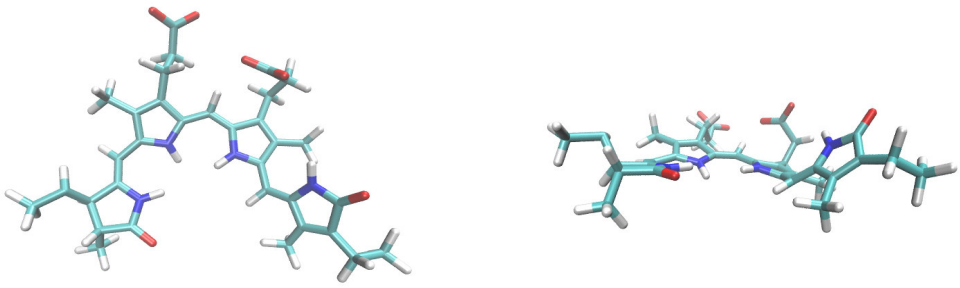

Figure 26. Structure of PCB after production run of 40 ns;
(left) on-top view; (right) side view

Instability of pyrrole-water during production run. Since both backbone of protein as well as chromophore remain stable during the production run, it could be estimated that also the water-network and especially the pyrrole-water should show only small fluctuations. For the purpose of monitoring the dynamics of the initial p-water (W1) the distances between the main hydrogen-bonding partners (N_δ of His260, backbone CO of Asp207, N_A, N_B, N_C and CO_A of PCB) and the oxygen of the p-water have been monitored, as depicted in Figures 27 and 28. Initially the water

stayed in close hydrogen-bonding distance to all partners but His260 for 10 ns. At around the 10th ns, fluctuations occur for all monitored distances, except between Asp207-CO and O of p-water. The water stayed in distances between 2 Å and 3 Å to Asp207 and between 3 Å and 4 Å to the cofactor for the next 10 ns. Afterward higher distances (up to 6 Å) can be observed between water and PCB. W1-PCB distances fluctuate from this timestep till the end of the simulation, at which they remain ~ 5.5 Å. An increased distance (up to 6 Å) can be observed for the Asp207-p-water distance after 36 ns of simulation time till the end. Thus the p-water is instable on its position between the four pyrrole-rings. After 35 ns of production time, the water flew out of its cavity (Figure 27 and 28) with increasing distances to all bonding partners up to 7 Å, with exception of His260. The water-molecule changes site of the PCB from below (near Asp207) to above PCB and stays near His260 (see Figure 29). Furthermore, the changes in the tiltangle of the pyrrole-water, with respect to the pyrrole-rings have been measured (not shown here). Constant changes of about 3° also identify the instability of named water.

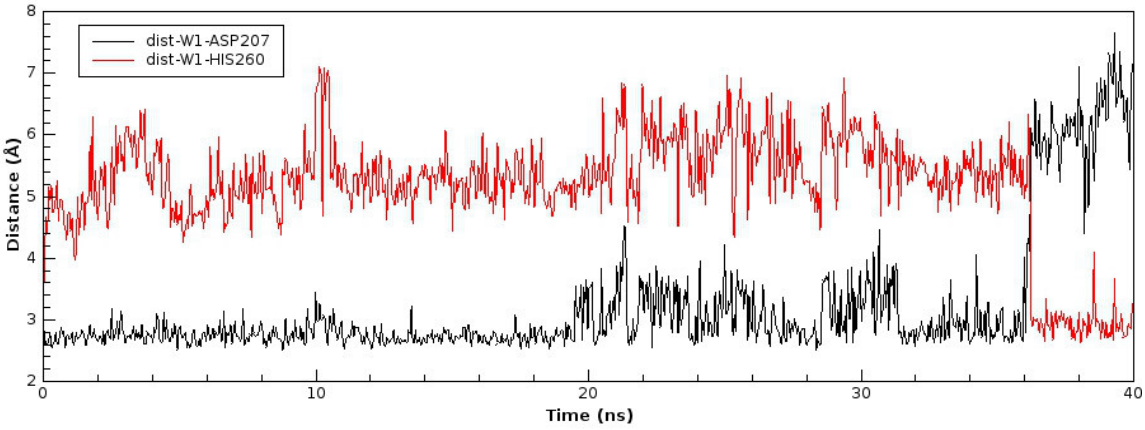

Figure 27. Evolution of distances between pyrrole-water W1, His260- N_δ and Asp207-backbone-CO during 40 ns classical of MM-MD

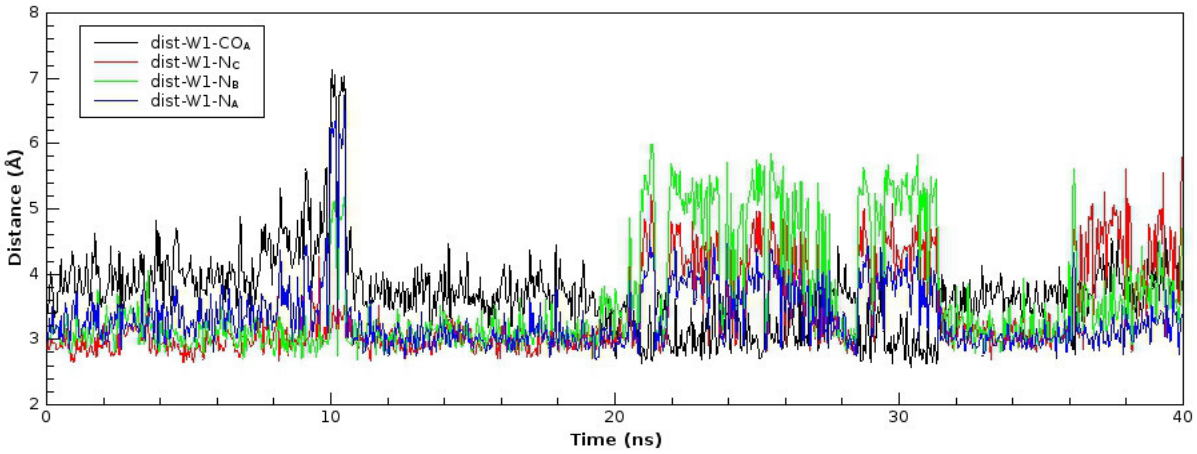

Figure 28. Evolution of distances between pyrrole-water W1, N_A, N_B, N_C and CO_A of PCB during 40 ns of classical MM-MD

Over the time-course of the MD-simulation, after the p-water left the position between the rings, different water molecules flew in and out of this position.

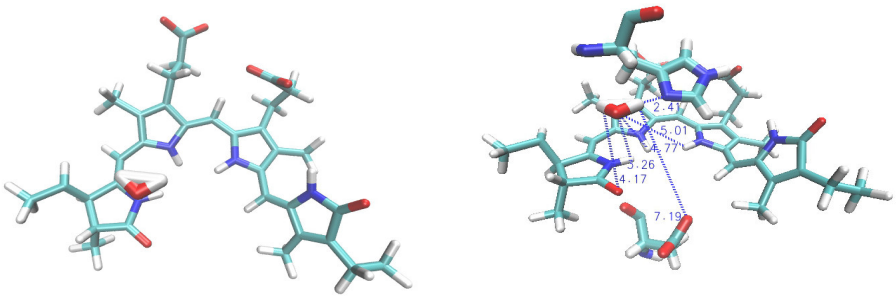

Figure 29. (left) Position of W8160 after 40 ns of production run; (right) with visualized distances to N_A, N_B, N_C, CO_A, His260-N and Asp207-O

Instability of solvent waters during production run. After the pyrrole-water (W1) flew away from its initial position at PCB it was replaced by a solvent water (W2) which again showed the unstable behavior of its predecessor. In Figures 30 and 31 the corresponding changes in distances between the oxygen of the solvent water W2 and N_δ of His248, backbone CO of Asp207 and N_A, N_B, N_C and CO_A of PCB are displayed. At around 8 ns of simulation time, the molecule dove into the chromophore binding pocket and stayed there in hydrogen-bonding distance (for at least 2 ns (see Figures 30 and 31) to named hydrogen-bonding partners (although W1 is still located between the pyrrole-rings at this time). Then it fluctuated for almost 20 ns until it again shifted onto the p-water location between 34 ns and 36 ns of production time. At this time W1 flew into the solvent. The lifetime of W2 on this position is only about 4 ns, then it also wanders back into the solvent, replaced by another one (not shown here).Observation of the changes in the tiltangle of W2 shows the same trend as noticed before with W1, namely highly deviating angles over the whole course of the simulation.

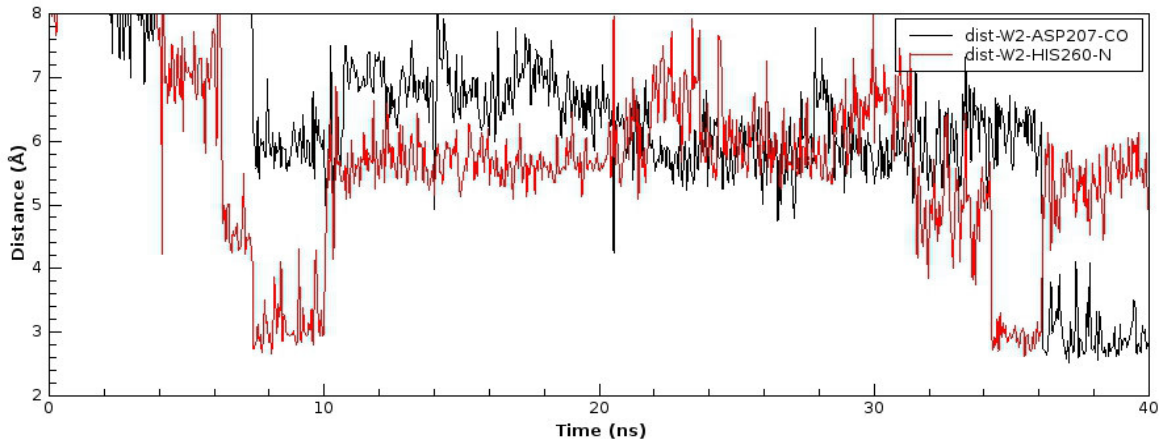

Figure 30. Evolution of distances between solvent water W2, His260-N_δ and Asp207-backbone-CO during 40 ns of classical MM-MD

It can be said conclusively that the initial water molecule at the p-water position is unstable, while solvent water molecules fly in and out of the cavity. Therefore, the unpublished findings of Daminelli et al. [64] are confirmed.

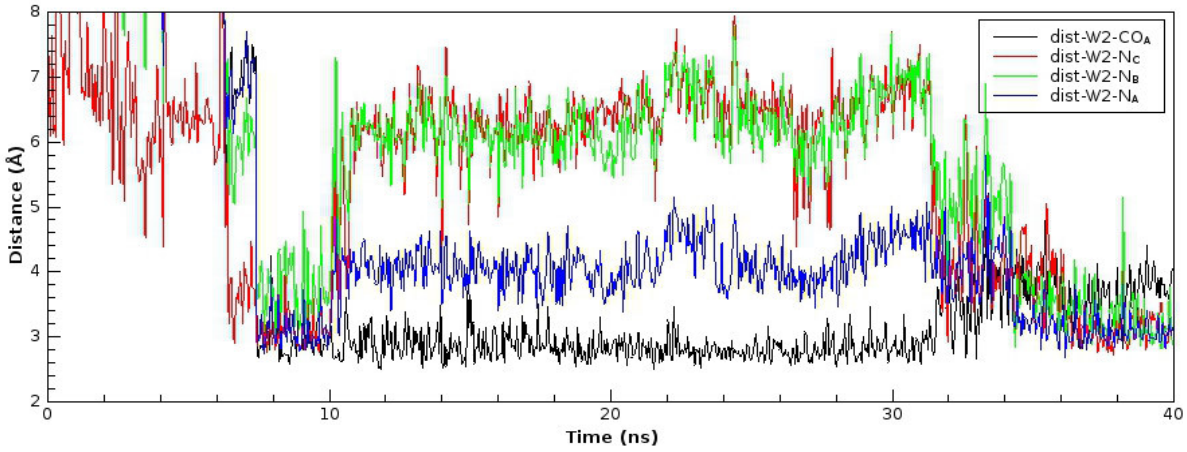

Figure 31. Evolution of distances between solvent water W2, N_A, N_B, N_C and CO_A of PCB during 40 ns of classical MM-MD

6.2. Computation of Mulliken charges and generation of a polarized force field

Mulliken charges have been calculated for all atoms (protein side-chains and waters) inside a 4 Å sphere centered at N_A of the PCB *via* a hybrid QM/MM approach, according to section 5.1.3. (for detailed information about the residues taken into account, please consult section 5.1.3. and Figure 20b). In this way, polarization effects of the cofactor on its environment and *vice versa* have been included into the force field. For a comparison between old point charges and computed Mulliken charges, please see Appendix H:

Of special interest is the effect of the protein environment on the charges of the chromophore. Bigger changes of about 0.04 – 0.3 can be noted for almost all PCB atoms. Especially the methylene, carboxylic and carbonyl groups, as well as the ring-nitrogens show high deviations from their original MM-charges. In contrast to this, the p-water shows only moderate changes of 0.05 for the oxygen and ~0.03 for the hydrogens. Protein side-chains inside the 4 Å sphere centered at PCB show the most deviations in near vicinity of the chromophore. Especially N and the corresponding H of His260 and His290, as well as C, CO and the aminofunction of Arg254 and C, CG and backbone O of Asp207 show higher deviations in their respective atom charges. In most cases already positive charges became even more positive and negative charges more negative. Thus, it can be seen that the chromophore is hugely affected by the charge environment of the protein. The amino acids have a big impact on the charges of PCB. In a reverse fashion, the chromophore affects the point charges of surrounding residues.

The “increased” positive or negative charges, especially of the ring-nitrogens of the chromophore might facilitate a more stabilized pyrrole-water on its position between the four pyrrole-rings. Therefore, the original MM-charges do in parts not correctly reflect the real charge environment.

These charges have then been incorporated into the existing force field and thus used to generate a polarized force field (polff). This has been used for subsequent MD-simulation (polff-MD), in which the stability of the pyrrole-water on its position between the four rings of PCB was subject of investigation.

6.3. Stabilization of pyrrole-water *via* polarized force field (polff-MD)

RMSD of Cph1 and chromophore PCB. The RMSD values for both protein and PCB have been computed with respect to the solvated X-ray structure [10] (see Figure 32): Compared to the MM-MD it is notable that the RMSD values for PCB are slightly lowered to an average RMSD of 1.2 Å and stay constant for the whole simulation. Also the RMSD's of the protein show a decrease in values (between 2 Å and 3 Å) and a smaller average RMSD of about 2.5-3 Å. Both decreases in RMSD may be attributed to a stabilization effect of the implemented polarized charges. Exceptions are the RMSD's of the protein between 30 ns and 40 ns, when an increase up to 7 Å can be noted. The reason for this are bigger movements of loop-regions and not a denaturation of the protein.

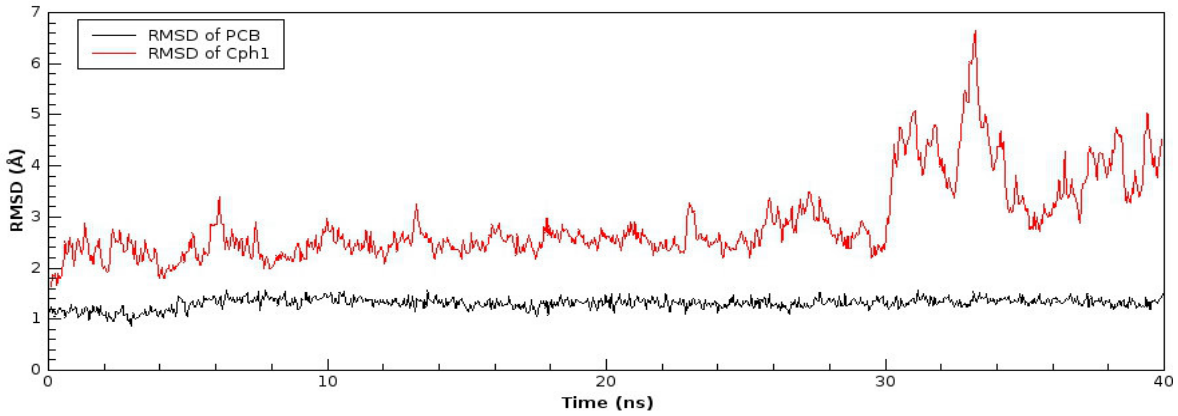

Figure 32. Average RMSD of Cph1 (red) and PCB (black) over time of 40 ns of production for MD with polarized ff

Structure and RMSF calculations of PCB. Comparing the structures of protein and chromophore from the polff-MD with the ones of the MM-MD-simulation (shown in Figure 33) reveals similarities in both. With the exception of loop regions, the protein structure from the polff-MD shows less deviations from the crystal structure than the one of the MM-MD (RMSD's of 3.5 Å and 4 Å, respectively). The biggest deviations between both structures can be found in the GAF region near the chromophore which may be due to the implementation of the polarized charges in this area. Furthermore, the structures of PCB are nearly identical, with only small differences in the tilting of the methyl groups of rings B and C (RMSD(MM): 1.25 Å; RMSD(polff): 1.2 Å). Also in the structure of the polff-MD, rings A, B and C are in plane (like it has been found in X-ray studies [10],[12]) with ring D out of the plane (same angle as in previous simulation).

The residues His260 and Asp207, which are essential for hydrogen-bonding of the p-water, are closer to the cofactor (as in MM-MD) and also the oxygens of Asp207 (both OH and CO) are directed towards the p-water position (in contrast to MM-MD, where both O-groups point slightly away from the p-water). Therefore, both residues are better available for H-bonding of the p-water.

RMSF calculations of PCB for the polff-MD reveal an identical pattern as for the MM-MD, but with strongly decreased RMSF values: RMSF of 6.9-7.5 Å for the biggest fluctuations. Similarly ring D (7.5 Å) and the corresponding bridge between rings C and D (RMSF=7.3 Å) are the most flexible parts, followed by the methyl groups of rings A, C and D (RMSF=7 Å) and the bridge between rings A and B (RMSF=6.9 Å).

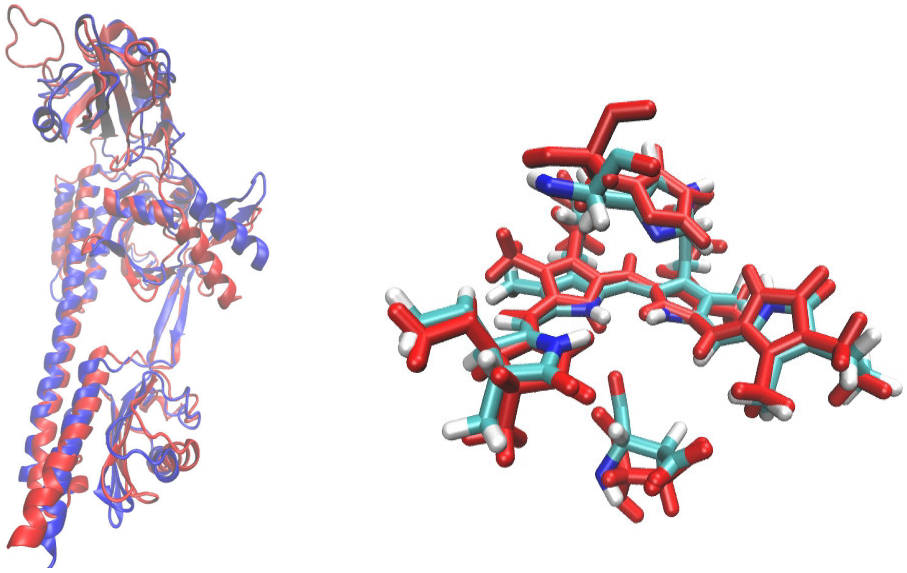

Figure 33. Structures of both Cph1 (left) sensoring domain; (right) PCB chromophore (with additional His260 and Asp207) for MD with polarized ff (blue and colored according to different atoms) compared to MM-MD (red)

It can thus be seen that the implementation of polarized charges diminish the fluctuations in the structure of the cofactor.

The stabilization effect of the polarized force field on the water-network inside the CBD in contrast to the standard force field can be seen, when looking at the dynamics of both crystal and solvent waters inside the CBD. For these purposes the distances between the six hydrogen-bonding partners (N_δ of His248, backbone-CO of Asp207, N_A, N_B, N_C and CO_A of PCB) and the crystal water W1, as well as the solvent water W3 have been monitored and are presented in the following:

Instability of pyrrole-water W1 during production run. W1 resides at the start at a position between the four pyrrole-rings of PCB (s. Figure 5), from which it flies away after only 1 ns of simulation time. Afterward it stays stable on a position underneath the chromophore and in close approximation to CO_A of PCB (O(PCB)-H(W1)-distance of about 3 Å) with only small fluctuations of less than 1 Å (not shown here).

Stabilization of solvent water W3 during production run. After 4 ns the empty cavity in the center of the pyrrole-rings of the cofactor is again occupied by a water molecule (W3) which moves inside from the solvent. Figures 34 and 35 display the continuous occupation of this molecule on the crystallographic pyrrole-water location [10] with only small fluctuations in the analyzed distances. It stays in close hydrogen-bonding distance (less than 3 Å) for all six binding-partners (N_δ of His248, backbone-CO of Asp207, N_A, N_B, N_C and CO_A of PCB) with exception of CO_A of PCB. The observed shifts in RMSD are all under 1 Å and thus tolerable. This indicates the desired stability of a water on the p-water position. Figure 36 displays the p-water after 40 ns of simulation time on its supposed position at PCB with the potential hydrogen-binding partners.

Tiltangles of W3 do not change (fluctuations are smaller than 1°) after 4 ns of simulation. Also it is notable that no additional water molecules are present at the position between the four pyrrole-rings of PCB (which is in contrast to the observations of the MM-MD).

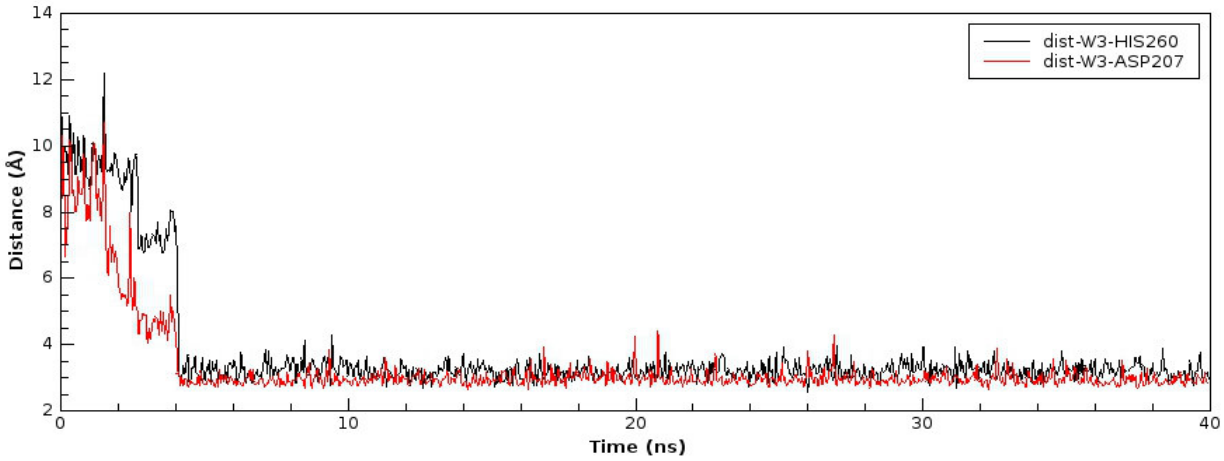

Figure 34. Evolution of distances between solvent water W3 and His260- N_δ and Asp207-backbone-CO during 40 ns of MD with polarized force field

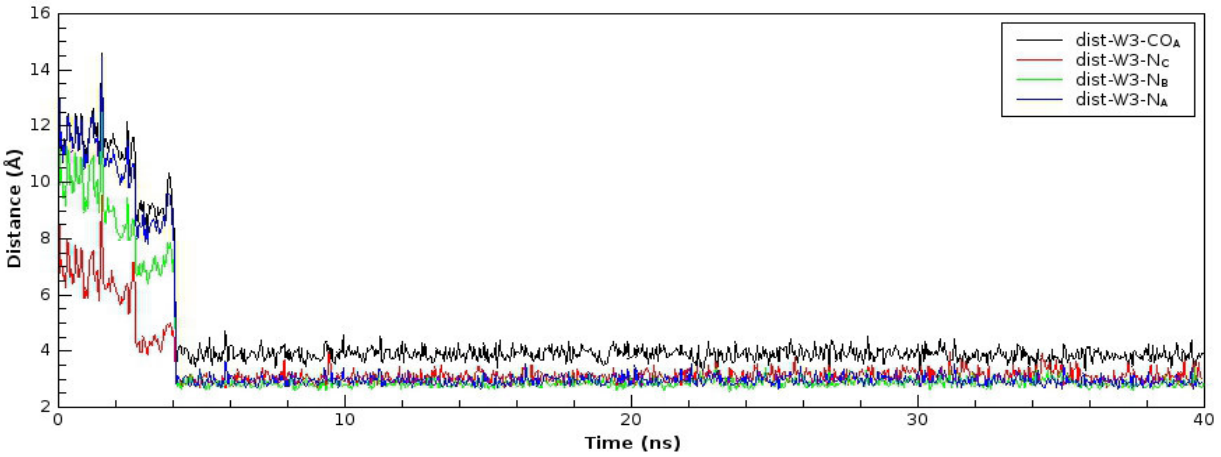

Figure 35. Evolution of distances between solvent water W3 and N_A, N_B, N_C and CO_A of PCB during 40 ns of MD with polarized force field

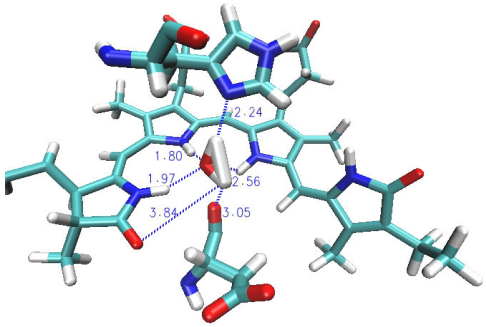

Figure 36. Position of W3 after 40 ns of production run with visualized distances to N_A, N_B, N_C, CO_A, N_δ of His260 and backbone-CO of Asp207

6.4. Discussion

In conclusion it becomes obvious that without polarized charges included into the force field, the waters, especially on the p-water location, are unstable and exchange very often during MD-simulations. In contrast to this, the addition of polarized charges for PCB and its surrounding residues leads to only one exchange in water molecules observed in the corresponding MD. The last water stays stable without bigger movements and fluctuations on its supposed position. Since only small differences between the structures of the chromophore in both MD simulations could be observed, the stabilizing effect is assigned do be mainly due to the usage of the polarized force field. This is due to the fact that the implementation of Mulliken charges incorporate polarization effects of the CBD-residues on the cofactor and *vice versa*. In classical MM-MD's these polarization effects are neglected, since the point-charges of protein residues have been determined for the Apo-protein (see section 3.1.). But the interaction of charges of cofactor and environment play important roles in stabilization of the surrounding water-network. The polarized charges help to reproduce the water-stability as observed in NMR [194],[195] and crystal studies [10]. Further repetitions of the named MD's (not shown here) show the same trend in stabilities and confirm the reproducibility of the results. Thus a polarized force field for at least a small region of interest (e.g. cofactor binding-site) can prove necessary, if MD-simulations show similar instabilities of water molecules. Since the computation of Mulliken charges, as well as the generation of a polarized force field can be time-consuming it should be tested beforehand if their application is necessary. In conclusion, polarized force fields provide a useful and effective way to generate a more realistic charge environment and overcome instabilities in MD-simulations. Additionally, it can help to improve computed Raman-spectra, as is discussed in the next chapter.

Chapter 7

QM/MM based Raman-spectra-calculation

Calculations of Raman-spectra for both MM-MD- and polff-MD-simulations have been carried out (as explained in section 5.1.4.). In the past, differences have been observed between computed Raman-spectra and experimental ones, especially regarding the NH in-plane rockings (broadening of corresponding band) [63]. It has been suggested that this may be due to the instability of the pyrrole-water on its position during normal MM-MD simulations. Therefore, it has been of interest, if spectra computed out of polff-trajectories will lead to an increased comparability to experimental data [165]. For this purpose QM/MM geometry optimizations have been done for 25 snapshots (of both MM-trajectory and polff-trajectory), using QM methods for the chromophore, the sulfide-link to the protein (including Cα of the corresponding Cys residue) and the p-water. The rest of the system has been treated with an MM force field. In the case of the MD simulation with polarized force field, the new charges have been added to the MM part of the optimization. Afterward the optimized structures were used for subsequent Raman-spectra computations according to procedure explained in section 5.1.4. Two different types of spectra are presented and discussed in the following: First, minimum energy spectra for both types of MD-simulations. For these, the snapshots with the lowest energy after optimization has been used, called minimum energy spectra (min-spectra). And second, sum-spectra were created by summing up of the data over all 25 snapshot-spectra. This is one way to include the influence of fluctuations in the protein environment and the stability of the p-water on the computed spectra (see section 4.4.4.). These computed Raman-spectra were compared to experimental Resonance Raman spectra of Cph1 by the group of Peter Hildebrandt [165].

7.1. Minimum energy Raman-spectra based on MM force field VS. polarized force field

In Figure 37a and 37b both calculated minimum energy spectra for MM-MD and polff-MD are depicted in comparison with the experimental Resonance Raman spectra [165]. The general trend is the same in both (MM and polff). The one based on the polff-MD shows differences in intensities and their distribution, compared to the spectrum of the MM-MD. The bands are more pronounced and have smaller bandwidth and are mostly of higher intensity in the former.

For discussion purposes, the complete spectra has been divided in three different regions. The low frequency region (600-1000 cm^{-1}), in which mostly out of plane movements (especially of the four rings) are present, the middle region (1000-1400 cm^{-1}) of single-bond vibrations and the so-called marker band region (1400-1700 cm^{-1}) which holds the C=C-stretchings of the methine (CH_2-) bridges and the NH-ip rockings (rockings) of rings B and C. The focus of the investigation has been on the marker band region and especially on the NH-ip bands. The spectra for the MM-MD will be referred to as MM-spectra and accordingly the spectra of the MD with polarized force field as polff-spectra. The results are shown in Figures 37a and 37b and Table 1.

Comparison of the low region for both spectra reveals certain differences, as for example the lower intensity (53) of the peak at 674 cm^{-1} in the polff case (80 in MM-spectra). Peaks at 727 and 762 cm^{-1} (MM) are downshifted compared to polff (713 and 754 cm^{-1}). Also the band at 769 cm^{-1} (MM) is upshifted. Most notable is for instance the splitting of a broad band (787 cm^{-1}) in the MM case into two distinct peaks (806 and 832 cm^{-1}) which create a doublet in the spectrum of the polff-MD.

Following an identical pattern, an additional doublet occur in the polff-spectrum at around 970 and 987 cm^{-1} which is not visible in the MM case. In general, the spectrum for the polarized trajectory shows more pronounced and clearly distinguishable peaks, where in the MM-spectrum often broad bands are visible. The region between 1000 and 1220 cm^{-1} shows no differences between both (with exception of an additional peak at 1150 cm^{-1} in polff-spectrum contributing to a doublet). In the region of 1300 – 1350 cm^{-1} (mostly CX_2 waggings and twistings) also strong deviations in intensities of a ratio of about 3:1 (polff:MM) between both computed spectra can be seen, accompanied by additional down- (1290 to 1279 cm^{-1} and 1307 to 1304 cm^{-1}) and upshifts (1320 to 1329 cm^{-1} and 1346 to 1347 cm^{-1}) in the polff-spectrum. Of great interest were the differences in the marker band region and the appearance of the NH-ip rocking bands. Therefore, a more detailed look on this region is provided in Figure 37b.

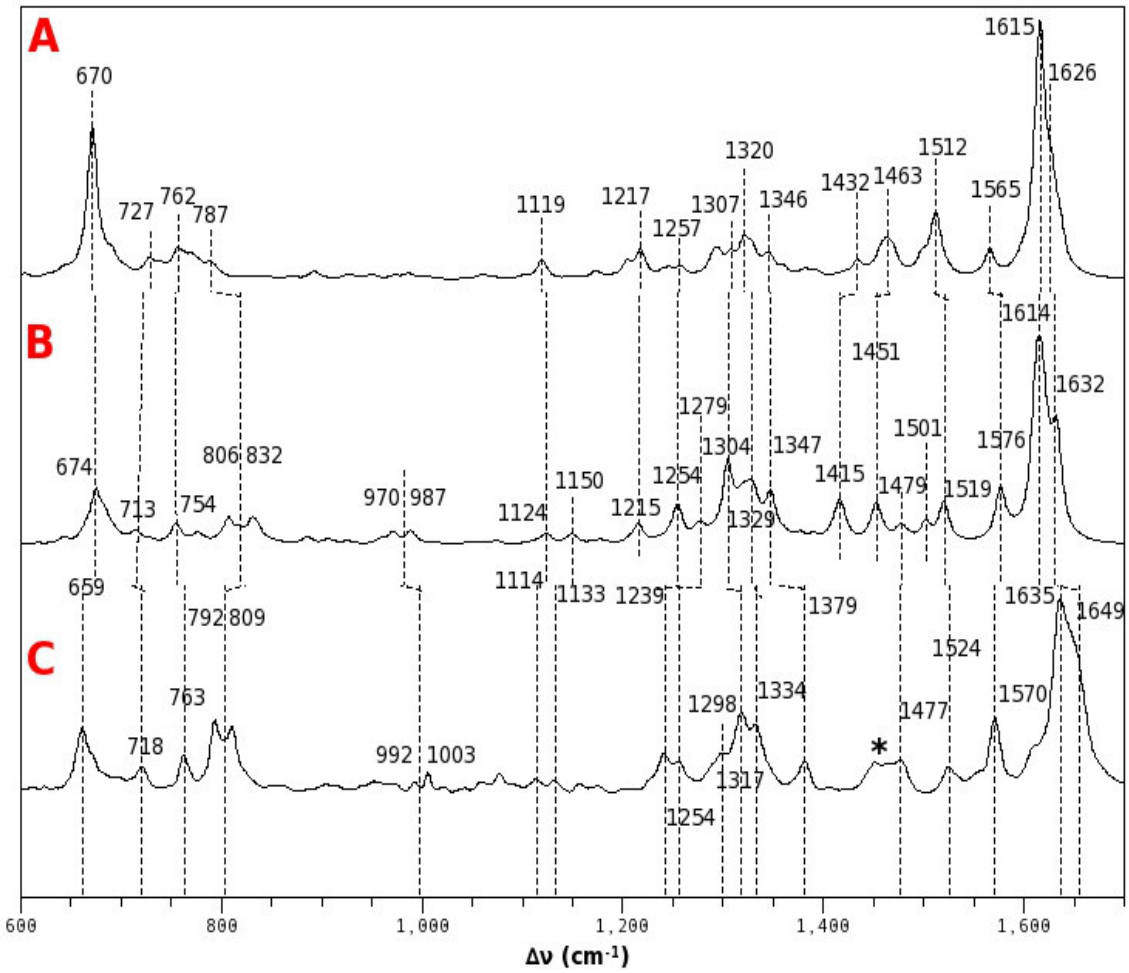

Figure 37a. Minimum energy spectra for both MM (A)- and polff (B)-MD simulations of Cph1 compared to experiment (C) Experimental Resonance Raman spectrum [165] (Cph1 in frozen solvent); asterisk in C denotes solvent bands

On first view, it can be seen that the peaks in the polff-spectrum are more pronounced and better distinguishable from each other. A downshift of 1415 and 1451 cm^{-1} compared to the MM-case (1432 and 1463 cm^{-1}) can be observed. These two bands correspond to the C=C (ring C and methine bridge between rings C and D) stretchings, as well as the NH-ip rockings of ring A in both-spectra or the NH-ip rocking of ring D in the case of the MM-spectrum. For a more detailed mode composition of the bands in the marker band region, take a look at Table 1. Furthermore, two

additional bands can be seen in the polff-spectrum at 1479 and 1501 cm^{-1} (C-C and C=C stretchings of ring B), where only a small shoulder at the peak at 1512 cm^{-1} is present in the MM-spectrum. The interesting NH-ip rockings of rings B and C (which consist mostly of these rocking/rocking movements) are visible in both spectra in different quality: The ones in the MM-spectrum at 1512 and 1565 cm^{-1} are downshifted in comparison to the polff-data (in which they are at 1519 and 1576 cm^{-1}). The band which in the past has been found mostly affected by the instability of the p-water in MD-simulations, is the one at 1565 cm^{-1} (MM) and 1576 cm^{-1} (polff) and corresponds mostly to the NH-ip rocking of rings B and C. In the MM-spectrum this band is flat and broad, while in the case of polff it is sharper and higher in intensity. Also the intensity-distribution between named two signals is weighted differently: In the case of MM, the band at 1512 cm^{-1} is higher than the one ate 1565 cm^{-1}. This is in contrast to the distribution for the polff-MD which is *vice versa*. Finally, the highest peaks in both spectra, are corresponding to C=C stretchings of methine bridges between rings C and D or B and C, as well, as O=C stretchings of rings A and D. They are at 1615 cm^{-1} (MM-spectrum) and 1614 cm^{-1} (polff-spectrum). The signal is of higher intensity (about two times) and the shoulder (corresponding to C=C stretchings of the methine bridge between rings A and B and rings B and C mostly) more visible in the latter spectrum.

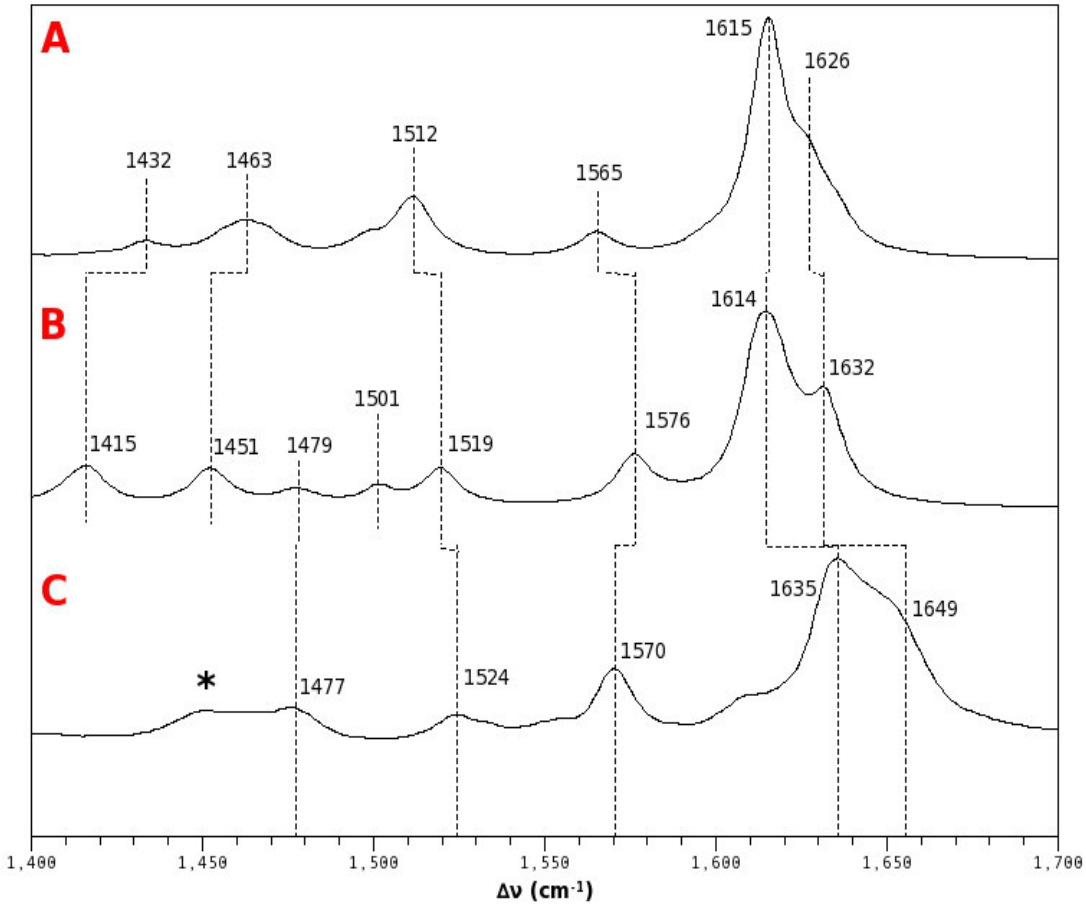

Figure 37b. Minimum energy spectra for both MM (A)- and polff (B)-MD simulations of Cph1, (C) Experimental Resonance Raman spectrum [165] (Cph1 in frozen solvent): zoom into the marker band region with visible NHip-rockings; asterisk in C denotes solvent bands

As a first conclusion, the usage of polff-MD's for spectra computations lead to spectra with sharper and more distinctive bands, especially the Nhip-rocking bands. This can be due to a more stable and stronger interaction between the cofactor and the pyrrole-water in the polff-MD.

Comparison of calculated spectra to the experimental Resonance Raman spectrum [165] (carried out in the group of Peter Hildebrandt) revealed the following: In the region of 600-1000 cm^{-1} there are some shifts in frequencies. For example both peaks at 670 cm^{-1} (MM) and 674 cm^{-1} (polff) are upshifted compared to the experiment (659 cm^{-1}). Also 713 cm^{-1} (polff) is downshifted, but closer to the experiment (718 cm^{-1}) than the corresponding peak in the MM-spectrum (727 cm^{-1}). The doublet at 792 and 809 cm^{-1} (exp) is only visible in the polff-spectrum (upshift of about 14 and 23 cm^{-1}) and missing in the MM case. The experimental bands at 992 and 1003 cm^{-1} are downshifted in the polff-spectrum of about 22 and 16 cm^{-1}, respectively. These are not visible in the MM data. Furthermore, in the middle-region the doublet at 1114 and 1133 cm^{-1} (exp) is only partly visible (only one signal at 1119 cm^{-1}) in MM, while in the polff-spectrum it is a complete doublet (upshifted by 10 and 17 cm^{-1}). Additionally, the two peaks at 1239 and 1254 cm^{-1} (exp) are sharper in the polff-spectrum (also upshifted about 15 and 25 cm-1). The peaks in the region between 1300 and 1350 cm^{-1} are sharper in the polff-spectrum and show the same intensity distribution as in the experiment. Further, the downshifts (13, 5 and 32 cm^{-1}, respectively) are smaller than in the case of MM.

ν/cm^{-1} (I) - MM-MD	**ν/cm^{-1} (I) - polff-MD**	**exp – Cph1-frozen solvent [165]**	**Normal Mode composition**
1432 (7)	1415 (24)	-	N-H rocking of ring D (MM), C=C stretching at ring C and methine bridge between rings C+D (polff)
1463 (6)	1451 (20)	-	C=O deformation + N-H rocking at ring A (MM+polff)
1512 (14)	1519 (31)	1524	N-H rocking at ring C + O=C stretching at ring B (MM), N-H rocking at rings B and C + C=C stretching of methine bridges between rings A+B and B+C (polff)
1565 (11)	1576 (40)	1570	N-H rocking at rings B and C (MM+polff), C=C stretching of methine bridge between rings B+C (polff)
1615 (100)	1614 (100)	1635	C=C stretching of methine bridge between rings C+D (MM+polff), O=C stretching at ring D (polff)
1626 (31)	1632 (75)	1649	C=C stretching (MM+polff), O=C stretching (polff)

Table 1. Vibrational frequencies of marker band region for both minimum energy spectra of MM- and polff-MD of Cph1 compared to experiment [165] and with assigned normal modes

Comparison of the marker band regions of both computed spectra to the experiment (see Figure 37b and Table 1.) first shows the missing of two signals in the experiment (at 1415 and 1451 cm^{-1} (polff) and 1432 and 1463 cm^{-1} (MM)), due to an overlap with solvent bands (denoted by an asterisk in Figure 37b) The next band at 1477 cm^{-1}(exp) is only present in the polff-spectrum (upshifted by 2 cm^{-1}). Furthermore, the band at 1501 cm^{-1} (polff), which is not visible in the MM-spectrum, can also not be seen in the experimental one. Both NH-ip rocking peaks at 1524 and 1570 cm^{-1} (exp) are downshifted in the MM case (12 and 5 cm^{-1}), while in the polff-spectrum they are only displaced by

5 and 6 cm^{-1}. Furthermore, the second band at 1570 cm^{-1} (exp) shows similarities to the corresponding one in the polff-spectrum by means of intensity and shape. Also the distribution of intensities between both NH-ip bands for polff (second higher than first) is in agreement to the experiment. Finally the peak at 1635 cm^{-1} and its shoulder at 1649 cm^{-1} (exp) is upshifted compared to both calculated spectra, to which the polff-spectrum is in better agreement.

Thus, the computed minimum energy spectrum for the MD with polarized force field is in better agreement to the experimental one in terms of vibrational frequencies and concerning intensities (and intensity-distributions). Further the usage of polff affects the shape of the NH-ip bands in such a way that they fit to the ones in the experiment better (more comparable frequencies and intensities) than without using polarized charges.

7.2. Comparison of both MM and polff methods including influences of protein and water fluctuations on the computed spectra

To visualize effects of protein and water fluctuations on the individual vibrational modes, sum-spectra have been computed (s. section 5.1.4.). In these the influence of the stabilization of the pyrrole-water due to the usage of polarized charges on the NH-ip rockings can be observed.

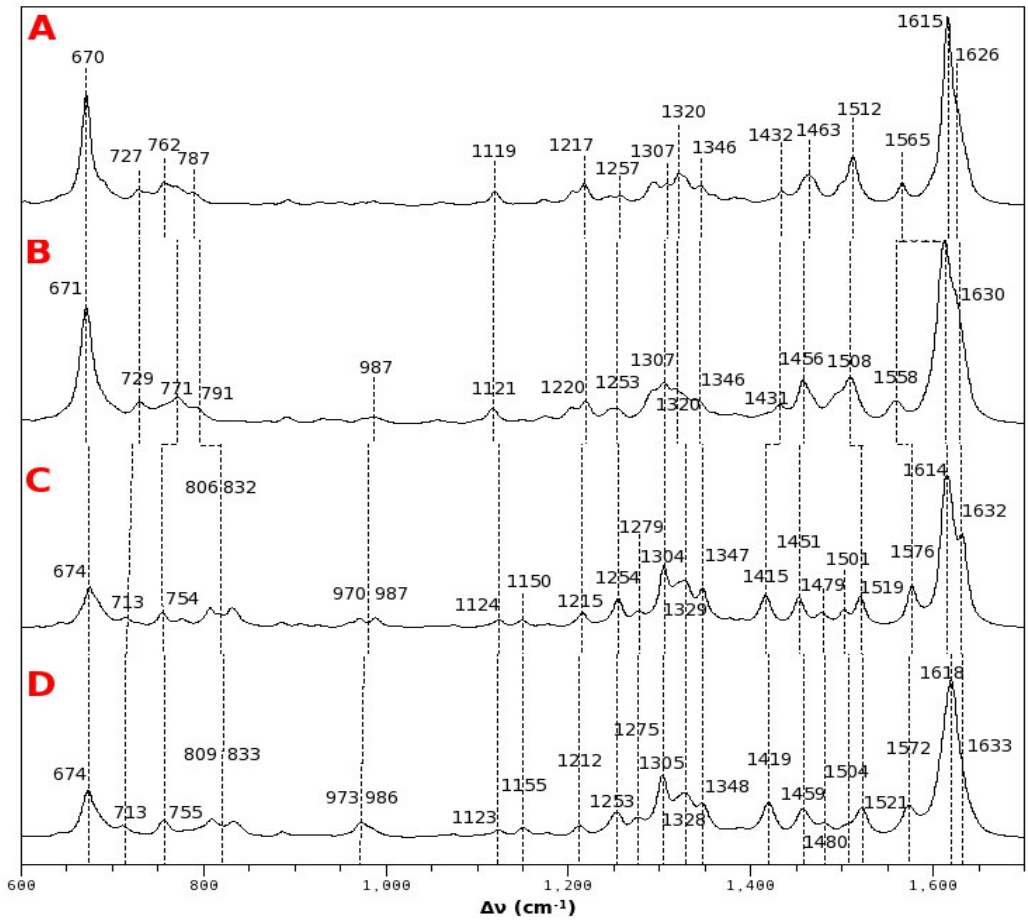

Figure 38. Comparison between calculated minimum energy spectra and sum-spectra of Cph1; (A) minimum energy spectrum of MM-MD; (B) sum-spectrum of MM-MD; (C) minimum energy spectrum of polff-MD; (D) sum-spectrum of polff-MD

In Figure 38 a comparison between both MM- and polff-sum- and minimum energy spectra is depicted. The differences in frequency-shifts (between minimum energy and the corresponding sum-spectra) are almost neglectable. Also the intensity distribution and the shape of the different bands remains the same as in the minimum energy spectra. The only notable differences are the appearance of an additional band at 987 cm^{-1} in the MM sum-spectrum, where a doublet is present in both experiment and polff case. The "triplet" at around 1307 cm^{-1} is less visible in the MM sum-spectrum, due to overlapping signals and shifts in frequencies of these, while the peak at 1456 cm^{-1} (MM) is sharper than in the respective minimum energy spectrum. Furthermore, the NH-ip rocking band at 1558 cm^{-1} (MM sum-spectrum) is broader and smaller and thus indicating the instability of the p-water and the shifts in the water-NH(PCB)-distances during the corresponding MD. In contrast to this, there are almost no changes from the minimum energy spectrum of polff compared to the respective sum-spectrum.

When comparing the signals' of the NH-ip rockings (Figure 38; peaks at 1558 cm^{-1} (MM) and 1572 cm^{-1} (polff)) of both sum-spectra, it can be seen that the one in polff is less wide, than the on in the MM sum-spectrum. Estimations of the standard deviation of the named Nh-ip signal in both MM and polff sum-spectra gives a ratio of about 5.3:3.6 (MM:polff). This demonstrates the decreased bandwidth in the polff sum-spectrum and thus indicates the stabilization of the p-water (and the water-NH(PCB)-interaction) with the application of a polarized force field even affects the computed Raman-spectra.

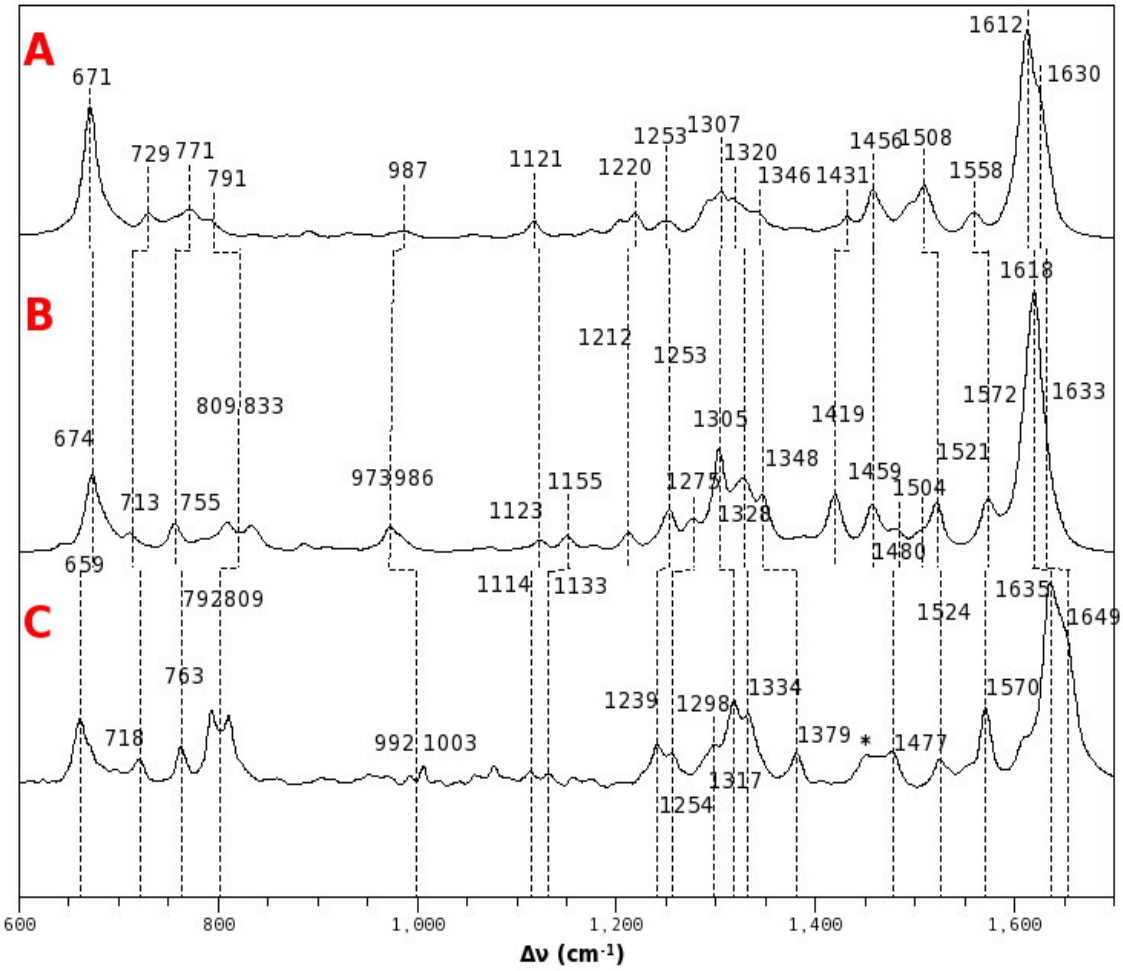

Figure 39a. Cph1 Sum-spectra for both MM (A)- and polff (B)-MD simulations, (C) Experimental Resonance Raman spectrum [165] (Cph1 in frozen solvent); asterisk in C denotes solvent bands

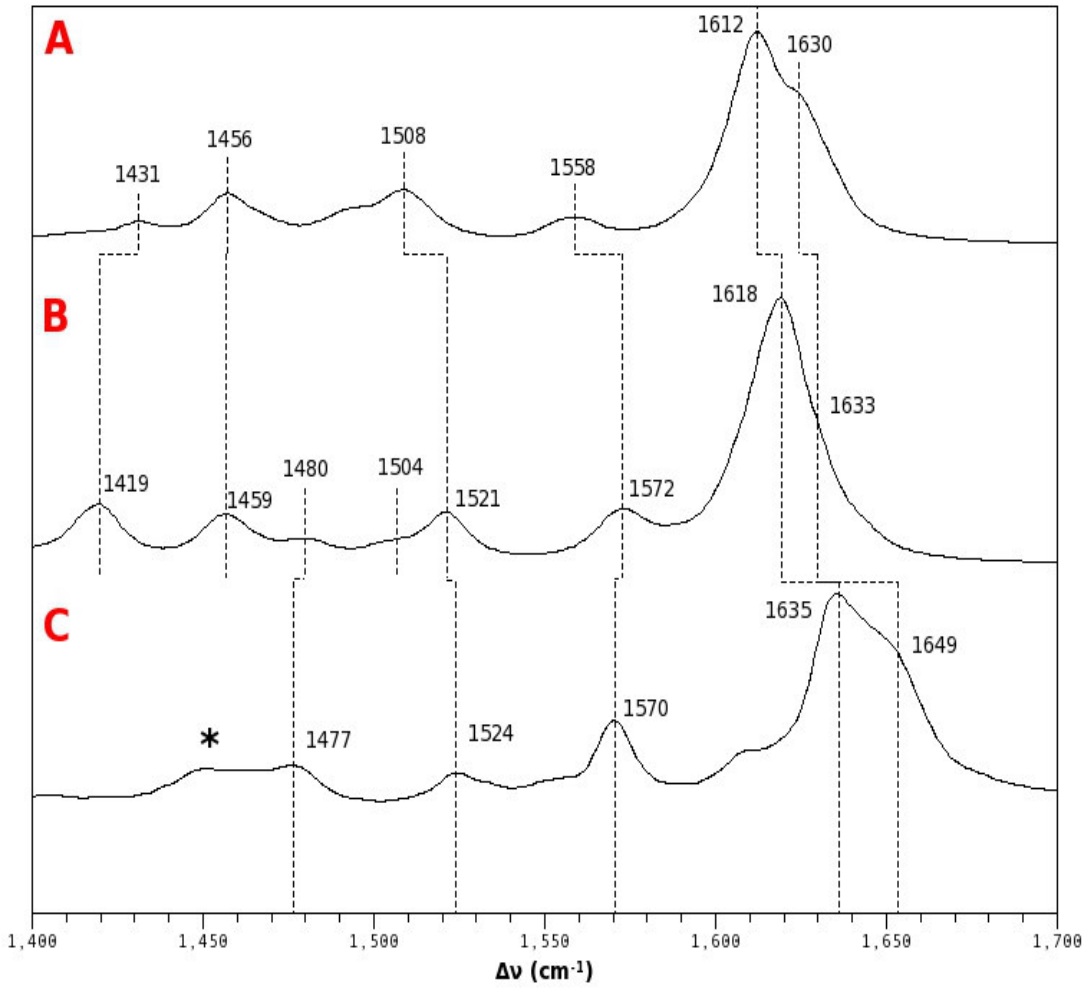

Figure 39b. Cph1 Sum-spectra for both MM (A)- and polff (B)-MD simulations, (C) Experimental Resonance Raman spectrum [165] (Cph1 in frozen solvent): zoom into the marker band region with visible NHip-bands; asterisk in C denotes solvent bands

The differences between calculated sum-spectra and the experimental Resonance Raman spectrum are the same as for the minimum energy spectra for both MM and polff (see Figure 39a). Therefore, only marker band regions of the sum-spectra were compared to the experiment (see Figure 39b and Table 2): The peaks at 1456 cm^{-1} (MM) and 1459 cm^{-1} (polff) are similar in frequency and intensity. The two bands at 1480 and 1504 cm^{-1} in the polff-spectrum are less visible than before (and 1480 cm^{-1} is 1 cm^{-1} more upshifted compared to the experiment than before). Both NH-ip rocking signals in the sum-spectra are still sharper for the polff-MD and here, they are in even better agreement to the experiment than in the minimum energy spectrum (also in terms of intensities). They are only shifted about 3 and 2 cm^{-1} (polff) from the experimental data, while in the case of the MM-sum-spectra the shifts in frequency are larger (13 and 14 cm^{-1}).

In this way, it has been demonstrated that the main goal of the reduction of bandwidth of the NH-ip rocking signals by stabilizing of the water network and especially the p-water with the help of a polarized force field has been successfully achieved.

ν/cm^{-1}- MM-MD	ν/cm^{-1}- polff-MD	exp – Cph1-frozen solvent [165]	Normal Mode Composition
1431	1419	-	N-H rocking at ring D (MM), C=C stretching at ring C and methine bridge between rings C+D (polff)
1456	1459	-	C=O deformation at ring A + N-H rocking at ring A (MM+polff)
1508	1521	1524	N-H rocking at ring C + O=C stretching at ring B (MM), N-H rocking at rings B and C + C=C stretching of methine bridges between rings A+B and B+C (polff)
1558	1572	1570	N-H rocking at rings B and C (MM+polff), C=C stretching of methine bridge between rings B+C (polff)
1612	1618	1635	C=C stretching of methine bridge between rings C+D (MM+polff), O=C stretching at ring D (polff)
1630	1633	1649	C=C stretching of methine bridges between rings A+B and B+C (MM+polff), O=C (stretching at ring A (polff)

Table 2. Vibrational frequencies of marker band region for both sum-spectra of MM- and polff-MD of CPh1 compared to experiment [165] and with assigned normal modes

7.3. Discussion

The minimum energy spectrum for the polff-MD shows sharper bands and intensity distribution patterns which are in better agreement to the experiment (especially in the marker band region, e.g. the NH-ip rocking signals) than the MM-min-spectrum. Also the frequencies of the individual vibrations in the polff-spectrum are more close to the experimental data. This is confirmed by the corresponding sum-spectra showing the same trend. The polff sum-spectra is in even better agreement to the experimental Resonance Raman data. This leads to the conclusion that the stabilization of the p-water (and the water network around PCB in general), due to the usage of polarized charges, directly affects the reproducibility of vibrational data with theoretical methods in a positive way.

Chapter 8

Summary and conclusions of part III

It is assumed that crystal waters are important for correct dynamics and processes inside protein environments (s. sections 1. and 3.1. for references). It is a major issue to simulate them correctly in theoretical applications. In preliminary work of Morginski et al. [63] and Daminelli et al. (unpublished) [64] an increased mobility of the water-network, especially the pyrrole-water, could be observed. This led to computed Raman-spectra which showed wider bands compared to experimental data. This was especially true for the very flat and broad signals for the the NH-ip rocking band for pyrrole-ring C. It was assumed that this could be directly related to the wrong treatment of water molecules with the applied force field. Currently available MM-MD methods might suffer from an insufficient treatment of the charge interactions between individual groups, especially in near vicinity of cofactors. This might lead to a destabilization of the water-network [57],[58],[59],[60],[61],[62]. The development of improved techniques for modeling of the charge-environment has thus proofed necessary in order to generate more realistic MD-trajectories [65], [66],[67].

In this work, the quantum-chemical computation of polarized charges for at least the chromophore PCB of Cph1 and the protein residues surrounding it, and its implementation in classical molecular mechanic force fields has been carried out. It has been shown in the case of Cyanobacterial Phytochrome Cph1 that these new charges stabilize the water-network in the vicinity of the cofactor. This is due to the more realistic treatment of polarization effects of cofactor and environment on each other.

Furthermore, these new polarized charges have been used within a QM/MM formalism for subsequent geometry optimizations of different snapshots and Raman-spectra calculations for both polff- and MM-trajectories. In this way, it has been demonstrated that the polff-MD leads to a better reproduction of experimental Resonance Raman results than spectra computed out of a classic MM-trajectory. Most significant are the changes in both intensities and appearances of the individual signals. The application of a polarized force field as done in this work, furthermore, leads to a more realistic band-shape of the NH-ip rocking mode, with the corresponding band becoming sharper and less wide (of about 1.7 cm^{-1}). This is in good agreement to the experimental data, also in terms of frequencies of these vibrational modes.

In that respect, this work demonstrates an approach for the generation of more accurate force fields for application with Cph1. A polarized force field reflects the interactions of partial charges in a more realistic way. The accuracy of QM/MM and Raman-spectra computations have been increased, helping to get a better reproduction of experimental data. Finally, the importance of the application of polarized charges into the existing force fields has been demonstrated. This method is helpful for systems which show similar problems with water instabilities and might stabilize the corresponding water-network. The more accurate treatment of charge distribution inside areas of interest of biological matter represent a valuable tool for prediction of structural, dynamic and spectroscopic properties and might help in the correct interpretation of experimental data.

Part IV.

Molecular dynamics and QM/MM based computations for phytochrome Agp2

Chapter 9

QM/MM calculations for determination of protonation state of two conserved histidine residues in near vicinity of the chromophore BV in Agp2

Agp2 is a bathy phytochrome with light-induced conversion from far-red light sensible parent state (Pfr) *via* Lumi-F and Meta-F intermediate states towards Pr by introduction of far-red light. From there on, the chromophore can either convert back to its parent state *via* red light treatment and the intermediate states Lumi-R and Meta-R or directly by a thermal back reaction from Pr to Pfr state. Herein the chromophore biliverdin (BV) is converting between different structures by rotation of ring D around the methyl bridge between rings C and D. The thereafter induced changes in the chromophore binding-domain (CBD) lead to a signaling cascade. The understanding of the triggering events of this signaling process are still not completely clear. Therefore, ongoing investigations are targeting the different states in the photocycle, while observing both chromophore and surrounding environment for apparent changes. The knowledge of each part of the mechanism and the contributing agents are of utter importance to get a complete picture of the photocycle.

As stated in section 3.2., the determination of the correct protonation of the two conserved histidines His248 and His278 in the vicinity of the cofactor is important, since they might play a significant role in the uptake and release of protons during the photocycle, as proposed for Cph1 [27]. The two protein residues and their protonation state might be of essence for the photoconversion: E.g. the thermal backreaction of Agp2 from Pr to Pfr *via* Keto-Enol-Tautomerie, as shown by Velazquez et al. [23] might be facilitated by (de-)protonation of His278. It is necessary for understanding and interpretation of experimental results and, furthermore, to achieve proper simulation conditions for theoretical studies to know the correct protonation states of these two histidines. For this purpose, nine different models (dd, de, dp, ed, ee, ep, pd, pe, pp) have been built as explained in sections 3.2. and 5.1.3., and treated with hybrid QM/MM methods in geometry optimizations in order to find the model which showed the least deviations from the crystallographic structure of Agp2 by Scheerer et al. [16]. Comparisons were always done to the protonated, solvated and water-equilibrated crystal structure [16] as described in section 5.2.1. Here, H278D accounts for protonation of N_δ-position, H278E means protonation of N_ε-postion and H278P means doubly protonated histidine. The results of these studies are described below.

General trends in the geometry optimizations of the nine different models. It can be observed that no change in the protonation states of both His and chromophore occurs during the geometry optimizations. Further it can be seen that the different protonation states of His248 and His278 have indeed an impact on the structural stability. In some models the changes are small, while in others (like the pe-model, as shown in 9.1.) they were strong. Also different amounts of alterations in the H-bond-network can be seen, depending on the used model. Furthermore, the potential bonds between N_δ or N_ε and their respective hydrogens of His248 and His278 are either elongating or shortening of around 0.003 A in the beginning of the optimization which then stayed stable till the end. As will be discussed later in more detail, there are some models, in which the distances do not stay steady, but instead fluctuate, indicating instabilities in the corresponding model. In the following the pe-model will be shown as an example for the instabilities introduced by His278 being protonated on its N_ε-position, followed by a demonstration of the more stable models with N_δ-protonation of His278.

9.1. His278 models with proton on N_ε-position

An example: pe-model. This model is a good example for the induced deviations from the crystal structure by a protonation of His278 on its N_ε-position.

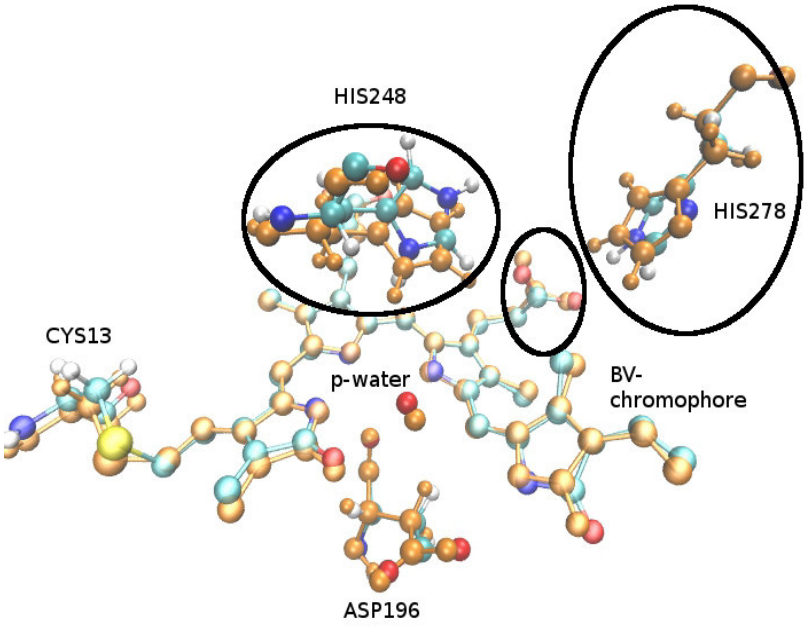

Figure 40. Optimized structure (colours) of pe-model compared to X-Ray [16] (orange), aligned at chromophore (NOTE: no protons at BV are shown; **BV is protonated at right PSC(C)-oxygen**)

In Figure 40 the optimized structure (in colors) is compared to the crystal structure (in blue). Deviations from the X-ray structure can be seen for Ser462, His248 and His278. Both histidines shift backwards from its original position, thus also affecting the surrounding water molecules which have to adjust to the altered environment. The chromophore itself shows only small differences to the crystallographic structure (e.g. a small twist of the PSC(C) due to the back shift of His278). These observations can also be found in the corresponding RMSD values (s. Table 3). The RMSD of BV is about 0.46 Å which indicates only small deviations. In contrast to this, both His248 and His278 give high values, 0.84 Å and 1.26 Å, respectively (see Table 3), suggesting that this model does not represent the native protonation state.

The distance-evolution of the N_ε-H-bond (s. Table 5) of His278 shows an oscillating nature, confirming the unstable structure of His278. In contrast, the analyzed N-H-bonding evolution of His248 shows only in the beginning a rise in the distance which then stays the same till the end of the trajectory. At the end of the optimization these bonds are in normal N-H-bonding range: His248-N_δ-H = 1.03 Å, His248-N_ε-H = 1.02 Å and His278-N_ε-H = 1.01 Å. Finally, the H-bond-network in near proximity to the histidines has been investigated (s. Table 4). It can be observed that no deviations occurred around His248 and near Ser462, while small ones are present near PSC of ring B (one additional H-bond of the oxygen of the carboxylic group to a water molecule). Contrary to this, the hydrogen-bonding-network connecting His278 to the cofactor decomposes completely which leaves His278 with no bond to the chromophore (also not *via* a water molecule).

All these deviations from the original crystal structure indicate that this model is not representing the native protonation state. The biggest differences concern His278 and adjacent parts of BV, as well as the H-bond connection between both. Such can be found for all models in which His278 was e- or p-protonated, leading to the suggestion that the models de, dp, ee, ep, pe, pp are incorrect.

9.2. Correctness of models with proton on N_δ-position of His278

An example: ed-model. The models with singly protonated His278 on the N_δ-position show less deviations from the crystal structure during the QM/MM optimization process, as shown for the ed-model: In Figure 41 the optimized structure (in colors) is compared to the crystal structure (in orange). Only small shifts can be observed, concerning His278, Ser462 and the PSC of ring C. This is also underlined by the corresponding RMSD values (s. Table 3) of BV, His248 and His278: 0.39 Å 0.41 Å and 0.78 Å, respectively. A fortiori, if compared to the ones of the pe-model. The NH-bond-evolutions for both histidines (s. Table 5) show stable bonds after an initial change. The absolute bond values for the NH-bonds at the end of the simulation are: His248-N_ε-H = 1.03 Å and His278-N_δ-H = 1.03 Å.

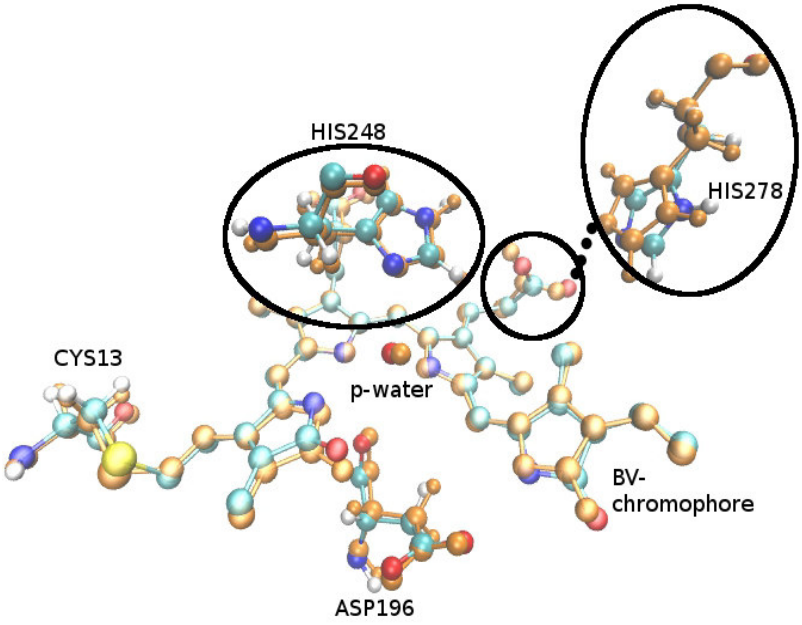

Figure 41. Optimized structure (colors) of ed-model compared to X-Ray [16] (orange); both aligned at chromophore (NOTE: no protons at BV are shown; **BV is protonated at right PSC(C)-oxygen**); dotted line indicate H-bond-interaction between PSC(C)-proton and free N_ε of His278

Furthermore, the qualitative investigation of the H-bond-network after the optimization shows no deviations from the network in the initial X-ray structure in the CBD. The network inside the computed QM-region stays the same during the optimization process (s. Table 4): Bonds from both H248-N's to the pyrrole-water and one water bridging to the PSC of ring C can be observed, as well as two bonds from His278-N_ε to the PSC of ring C of the cofactor, directly and *via* water-bridge. This is indicating structural stability of this model. These trends occur in all H278D models and indicate a correctness of the dd-, ed- and pd-models in contrast to the six others. In the case of the ed-model these findings were even further established by computations with larger basis set (6-31+G*) for all hydrogen atoms. Here, the RMSD values are even lower (BV: 0.40 Å, H248: 0.24 Å, H278: 0.71 Å), demonstrating even less deviations from the initial crystal structure.

9.3. Determination of correct protonation-model

These findings of models with single protonated His278 on its N_δ showing the least deviations from the initial crystal model, lead to the conclusion that one of the three models dd, ed or pd has to be the correct one. Therefore, all computed RMSD values (Table 3), stabilities of H-bond-networks (Table 4) and stabilities of N-H-bonds (Table 5) for all nine models are compared to each other.

RMSD (heavy atoms)	**H248D H278D**		**H248D H278E**	**H248D H278P**
BV	0.49		0.52	0.50
H248D	0.40		0.56	0.53
H278	0.81		1.07	0.90
RMSD (heavy atoms)	**H248E H278D**	**H248E* H278D***	**H248E H278E**	**H248E H278P**
BV	**0.39**	**0.40**	0.46	0.44
H248E	**0.41**	**0.24**	0.63	0.64
H278	**0.78**	**0.71**	0.72	0.84
RMSD (heavy atoms)	**H248P H278D**	**H248P* H278D***	**H248P H278E**	**H248P H278P**
BV	0.40	0.49	0.46	0.45
H248P	0.30	0.30	0.84	0.44
H278	0.87	0.95	1.26	0.70

Table 3. Computed RMSD values for all nine QM/MM geometry optimized models with aligned chromophore and with respect to the solvated + water-equilibrated crystal structure [16]; asterisk denotes the models computed with large basis set for hydrogen atoms (6-31+G*)

When the computed RMSD values of all models are compared to each other, the ones with single protonation of N_δ of H278 yield the smallest values. E.g. dd (RMSD(His248): 0.40 Å, RMSD(His278): 0.81 Å), de (RMSD(His248): 0.56 Å, RMSD(His278): 1.07 Å) and dp (RMSD(His248): 0.53 Å, RMSD(His278): 0.90 Å) compared to each other show the lowest RMSD's for the dd-model (differences of about 0.16 Å/0.13 Å for His248 and 0.26 Å/0.09 Å when comparing dd to de/dp). The same is true for the ed- (RMSD(His248): 0.41 Å, RMSD(His278): 0.78 Å), ee- (RMSD(His248): 0.63 Å, RMSD(His278): 0.72 Å) and ep-models (RMSD(His248): 0.64 Å, RMSD(His278): 0.84 Å), with the exception of the RMSD for His278 in ed which is 0.06 Å higher than in the ee model. But since also the RMSD for BV (0.39 Å) is smaller in ed than in ee (0.46 Å) it does not account to much. Here, ed shows differences to ee/ep of about: 0.07 Å /0.05 Å for BV, 0.22 Å/0.23 Å for His248 and -0.07 Å/0.06 Å for His278. Similar results can be seen for the pd-, pe- and pp-models, in which the RMSD of BV is 0.06 Å, for His248 about 0.54 Å and for His278 about 0.39 Å smaller when comparing pd to pe and about 0.05 Å for BV, 0.14 Å for His248 and -0.17 Å for His278 when comparing pd to pp. Here, pd gives a higher RMSD for His278 than in pp, but the other values are still lower.

The lowest RMSD's are always given for the H278D-models (the above mentioned exceptions can be neglected as shown below). Comparison of dd, ed and pd to each other in terms of structural stability reveals similarities in RMSD's which could mean a less important role of the protonation state of His248 for structural stability during the QM/MM optimizations. The dd-model shows slightly higher values for BV and His278 (0.1 Å/0.09 Å for BV and -0.01 Å/0.1 Å for His248 and 0.03 Å/-0.06 Å for HI278 when comparing to ed/pd), whereas the ed-and pd-models were alike in terms of RMSD. The ed-model has a smaller RMSD for His278 (0.09 Å) and for BV (0.01 Å) as in the pd-model. The RMSD for H248 is smaller in the pd-model (0.11 Å). ed shows smaller values for BV and His278 compared to both dd and pd. Simultaneously, pd has smaller RMSD values for BV and His248 compared to dd. Therefore, either ed or pd represents the correct protonation state.

With the help of further computations with larger basis set for hydrogens (6-31+G* for H's; see section 5.2.4.), it can be further differentiated between ed and pd (s. Table 3): Here, the ed-model shows the lowest RMSD values in all three investigated structural motifs (BV: 0.09 Å, His248: 0.06 Å, His278: 0.24 Å) compared to pd and thus makes it the more probable protonation model.

	H248D H278D		**H248D H278E**	**H248D H278P**
Stable/not stable	✔/✕		✕ (H278)	✕ (H278)
	H248E H278D	**H248E* H278D***	**H248E H278E**	**H248E H278P**
Stable/not stable	✔	✔	✔/✕ (H278)	✕ (H278)
	H248P H278D	**H248P* H278D***	**H248P H278E**	**H248P H278P**
Stable/not stable	✔/✕ **(H248)**	✔/✕ **(H248/H278)**	✕ (H278)	✕ (H278)

Table 4. Comparison of H-bond-network stability in all nine QM/MM geometry optimized models with respect to the solvated + water-equilibrated crystal structure [16] (✔: stable network → no changes; ✕: unstable network → many + crucial changes, concerning named residue; ✔/✕: moderate changes not affecting general stability → e.g. exchange of bonding partners); asterisk denotes the models computed with large basis set for hydrogen atoms (6-31+G*); denoted in parentheses are the residues which experience the biggest changes of the corresponding H-bond-network

In Table 4 the different stabilities of hydrogen-bond-networks of all nine models are compared to each other. It can be seen that in models, where His278 is e- or p-protonated, this resulted in the loss of most or all H-bonds from H278 to the chromophore. And also in the case of the H278D models only in the ed model a complete unchanged network can be observed at the end of the QM/MM optimization. In the pd-model additional H-bonds concerning His248 are observed. Additionally, in case of the corresponding calculations with bigger basis for H's, His278 loses H-bonds towards BV and surrounding water molecules.

In Table 5 all N-H-bond-evolutions of the two histidines during the QM/MM optimization process are monitored, whereas strong fluctuations indicate unstable bonds and thus unstable structures, i.e. less correct models. With the exception of ed-and ep-models, major deviations in the bond-evolutions of all other formations are always been observed. Especially models with N_ε-protonation of His278 yield fluctuating NH-bonds, with the exception of the ep-model. But the results concerning RMSD's and H-bond-network stabilities demonstrate the ep-model to be less probable to be the correct one. Additionally, in the case of pd, it should be mentioned that the N_ε-H-bond of His248 was subject to severe variations and is very long (1.05 Å) in the QM/MM optimized structure which shows the trend of this proton to to undergo a translocation to the PSC of ring C (observed for both applied basis sets for hydrogens). This renders the pd-model more unlikely to be the correct one. Further, this confirms that the ed-model yields the most stable structure after QM/MM optimization, with both N-H-bonds increasing/decreasing only in the beginning and then staying stable.

bond between N_δ or N_ε and H of His	H248D H278D		H248D H278E	H248D H278P
H248	stable		stable	stable
H278	unstable		unstable	δ:stable | ε:unstable
	H248E H278D	**H248E* H278D***	**H248E H278E**	**H248E H278P**
H248	**stable**	**stable**	unstable	stable
H278	**stable**	**stable**	unstable	both:stable
	H248P H278D	**H248P* H278D***	**H248P H278E**	**H248P H278P**
H248	δ:stable | ε:unstable	both:unstable	both:stable	both:stable
H278	unstable	unstable	unstable	δ:unstable | ε:stable

Table 5. N-H-bond evolution/stability of His248/His278 in all nine optimized models during QM/MM geometry optimization with respect to the solvated + water-equilibrated crystal structure [16]; asterisk denotes the models computed with large basis set for hydrogen atoms (6-31+G*); "stable" means constant bond-length during the length of the optimization, while "unstable" denotes strong fluctuations in the respective bonds during optimization procedure

In conclusion, the overall structure of the chromophore and its surrounding protein and water environment is affected by the protonation state of the two investigated histidines. His278 is most likely N_δ-protonated due to discussed RMSD-values, H-bond-network stability and NH-bond-evolution and His248 is N_ε-protonated. Thus, the most probable and correct protonation state for His248 and His278 will be the ed-model. This results fit nicely with recent finding of Velazquez et al. [23] who predicted the thermal degradation route from Pr to Pfr state *via* a Keto-Enol equilibrium, in which His278 is probable protonated on the N_ε-position by the proton from the PSC of ring C (as observed *via* RR). It would make a singly protonated His278 on the N_δ-position in the Pfr-state necessary to yield a free N_ε which then can be protonated in the Pr-state.

9.4. Models with deprotonated PSC(C)

General trends. Additional QM/MM calculations of models (dd, de, dp, ed, ee, ep, pd, pe, pp) with deprotonated PSC(C) have been carried out (see section 5.1.3.). The reason for this has been the investigation of a possible H-translocation from N_ε of His278 (for H278E/P models) to the PSC(C) (as it might occur in Pr-Pfr-conversion [23]). This would proof a protonated PSC(C) in the Pfr state of Agp2. Therefore, all computed results were compared to the initial crystal structure [16].

Due to high RMSD values (s. Table 6), H-bond-network instabilities (s. Table 7), as well as deviations in the corresponding N-H-bonds (s. Table 8) a high instability in all H278D models is revealed. The models with the least deviations from the initial crystal structure are the ones with p-protonation of His278. In all models containing a N_ε-H on the His278, no proton transfer can be observed, but still a strong tendency towards it, as shown in the following for the dp-model.

The dp-model will be discussed in more detail as an example. In Figure 42 it can be observed that the optimized structure of the chromophore (in colors), as well as the water molecules and both His in the QM/MM optimized dp-model are in good agreement to the crystal structure (in orange). This is further underlined by small RMSD values for BV, His248 and His278: 0.48 Å, 0.57 Å and 0.65 Å, respectively. The biggest structural change occur for the p-water which shifts away from BV. The hydrogen-bonding-network shows only very small deviations in the region of His278, where one (PSC(C))O-(His278)-N_{ε}-H bond is lost. The N-H-bond-lengths vary only in the beginning of the optimization procedure and stay stable afterward (except the N_{ε}-H-bond; see next paragraph).

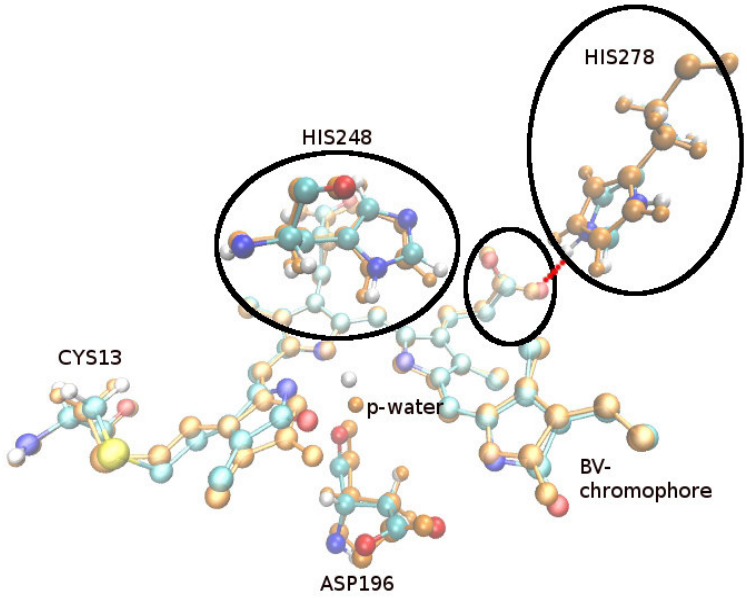

Figure 42. QM/MM geometry optimized structure (colors) of dp-model compared to X-Ray [16] (orange); both aligned at BV (NOTE: no protons at BV are shown; **BV is NOT protonated at PSC(C)**); dotted red line indicate hydrogen-bond-interaction between free oxygen of PSC(C) proton at ε-nitrogen of His278

Notable is the behavior of the His278-N_{ε}-H connection to the PSC(C) oxygen. A strong decrease in this distance can be observed in all His278E/P models. The N_{ε}-H-bond increases (up to 1.05 Å in the end) while simultaneously His278 shifts nearer to the PSC(C). Although the distance between His278 and PSC(C) always stays over 1.7 Å, this behavior indicates a strong tendency of the proton to leave His278 and switch to the PSC(C).

Comparison of different protonation models lead to confirmation of protonated PSC(C) (BV). All computed models are analyzed in terms of RMSD's, stability of H-bond-networks and N-H-bonds. In Table 6 to 8 the corresponding results are presented. The RMSD's for His278 are generally higher than with a protonated PSC(C). Furthermore, the highest deviations from the crystal structure occur for the models with His278D protonation and the lowest with the His278P protonation. The dp-model shows differences in RMSD of about -0.05 Å/-0.09 Å for BV and -0.31 Å/-0.07 Å for His248 and 0.57 Å/0.26 Å for His278 (compared to dd/de). Further, ep shows differences in RMSD of about 0.03 Å/-0.08 Å for BV and 0.18 Å/0.11 Å for His248 and -0.02 Å/-0.07 Å for His278 (compared to ed/ee). The pp-model finally shows RMSD-differences of about 0.2 Å/0.06 Å for BV and 0.18 Å/0.21 Å for His248 and 0.14 Å/0.29 Å for His278 (compared to pd/pe). So the best choice will be either dp, ep or pp. It is difficult to further distinguish between these three. Nevertheless since ep and pp have lower RMSD values for BV and His248, they seem more plausible.

RMSD (heavy atoms)	**H248D H278D**	**H248D H278E**	**H248D H278P**
BV	0.43	0.39	0.48
H248D	0.26	0.50	0.57
H290	1.22	0.91	0.65
RMSD (heavy atoms)	**H248E H278D**	**H248E H278E**	**H248E H278P**
BV	0.41	0.36	**0.44**
H248E	0.51	0.44	**0.33**
H278	0.86	0.81	**0.88**
RMSD (heavy atoms)	**H248P H278D**	**H248P H278E**	**H248P H278P**
BV	0.60	0.46	**0.40**
H248P	0.54	0.57	**0.36**
H278	0.92	1.09	**0.78**

Table 6. Computed RMSD values for all nine QM/MM geometry optimized models with aligned chromophore and respect to the solvated + water-equilibrated crystal structure [16]

The qualitative investigation of H-bond-network stabilities, which is shown in Table 7, helps further differentiating between these three models. From the three chosen models, the ep-model has the least deviations from the initial H-bond-network as observed in the crystal structure and thus supports the estimation that this represents the most probable model for a deprotonated PSC(C). In the case of dp and pp, there are always more changes in H-bonding for at least one of the two histidines: In dp the network changes around His278 moderately, while in pp it changes around His248 strongly.

	H248D H278D	**H248D H278E**	**H248D H278P**
Stable/not stable	✕ (H278)	✔/✕	✔/✕ (H278)
	H248E H278D	**H248E H278E**	**H248E H278P**
Stable/not stable	✕ (H278)	✔/✕ (H278)	✔/✕
	H248P H278D	**H248P H278E**	**H248P H278P**
Stable/not stable	✔/✕ (H248)	✔	✕ (H248)

Table 7. Comparison of H-bond-network stability in all nine QM/MM geometry optimized models with respect to the solvated + water-equilibrated crystal structure [16] (✔: stable work → no changes; ✕: unstable network → many + crucial changes, concerning named residue; ✔/✕: moderate changes → e.g. exchange of bonding partners); denoted in parentheses are the residues which experience the biggest corresponding changes of the H-bond-network

Finally, the only model with no fluctuations in the N-H-bond-evolutions for both His248 and His278 is the ep-model (as shown in Table 8), thus rendering this the one to be in the best agreement with the X-ray structure and the correct one for a situation with a deprotonated PSC(C) at BV.

bond between N_δ or N_ε and H of His	**H248D H278D**	**H248D H278E**	**H248D H278P**
H248	stable	stable	stable
H278	unstable	unstable	δ:stable |ε:unstable
	H248E H278D	**H248E H278E**	**H248E H278P**
H248	unstable	unstable	**stable**
H278	stable	stable	**stable**
	H248P H278D	**H248P H278E**	**H248P H278P**
H248	stable	stable	stable
H278	unstable	unstable	δ:stable |ε:unstable

Table 8. N-H-bond evolution/stability of His248/His278 in all nine optimized models during QM/MM optimization with respect to the solvated + water-equilibrated crystal structure [16]; asterisk denotes the models computed with large basis set for hydrogen atoms (6-31+G*); "stable" means constant bond-length during the length of the optimization, while "unstable" denotes strong fluctuations in the respective bonds during optimization procedure

Conclusively, the indicated tendency of a proton transfer from N_ε of H278 to BV confirms the findings of the investigations of the CBD with a protonated PSC(C): If the proton on N_ε of His278 in the ep-model would completely leave its position and change to the propionate of ring C of BV, this would lead back to the ed-model with protonated PSC(C). This is in agreement with the results presented in 9.3., in which the ed-model is stated as the model with the least deviations from the crystal structure and thus the correct one. These findings underline Resonance Raman findings that the PSC of ring C is protonated in its native Pfr state [23]. Furthermore, this confirms His278 being protonated on N_δ, while His248 is protonated on N_ε.

9.5. Discussion

In summary the comparison of QM/MM optimized models accounting for different protonation states for the two histidines His248 and His278 in near vicinity of BV (protonated PSC(C)) with the respective crystal structures revealed the ed-model (His248: protonation at N_ε; His278: protonation at N_δ) as the most probable native protonation state. Further computations and analysis of the respective models with deprotonated PSC(C) confirmed these findings. Also, the presence of a proton on PSC(C) as already suggested by RR-experiments [23] could be confirmed.

Similar studies have been recently published, concerning the protonation state of the two conserved histidine residues in both Cph1 [27] (His260, His290) and Agp1 [28] (His20, His280) in their parent Pr-states:

Cph1 has been found to have two different protonation patterns of His260 and His290 which are attributed to Pr-I- and Pr-II-states and thus refer to a ph-dependent heterogeneity [27]. Both differ in the protonation of His260. Pr-I resembles a doubly protonated and cationic His260, while His290 is protonated at the N_ε. In contrast to this, Pr-II shows single N_ε-protonation of both His' (pe- and ee-models, respectively). Velazquez et al. [27] suggested His260 as a proton-sink and -source and essential for the conformational heterogeneity of the CBD.

Takiden et al. [28] investigated the protonation pattern of His250 and His280 in Agp1 and determined also two possible protonation states [28]: Either both His' carrying a proton on N_ε, or His250 is protonated on N_ε, while His280 is charged and carries two protons (ee- or ep-models, respectively). Again a heterogeneity of the Pr-state might be possible [28].

The findings concerning Agp1 are in good agreement to the here presented results of Agp2. Here, a homogeneous Pfr-state with ed-protonation of His248 and His278 has been proposed. Velazquez et al. [23] suggested the release of the proton from PSC(C) and subsequent uptaking by His278 during Pfr-Pr-conversion which would lead to a negatively charged propionate at ring C and a doubly protonated and a positively charged His278. This is exactly what Takiden et al. [28] observed in the Pr-state of Agp1: an ep-protonation state of the two histidines and a deprotonated PSC(C) at BV. Furthermore, studies with a deprotonated PSC(C) showed an ep-model to be the most promising one for the correct protonation of the two histidines. This is also in agreement to the findings of Takiden et al. [28]. Conclusively, there might be a resemblance between the Agp1 (Pr state) findings and the ones for Agp2 (Pfr state) in terms of proton-translocations between PSC(C) and His278 and protonation states of the two histidines in general. This could mean a shared and similar photocycle with similar characteristics in both phytochromes.

In contrast to this, His290 in Cph1 carries in both parent Pr-states only one proton at N_ε [27]. This leads to the suggestion that both Cph1 and Agp1/Agp2 follow different photoconversion mechanisms, concerning the role of their respective His290 and His280/His278. A common feature of all three systems is the protonation of N_ε in His260 in Cph1, His250 in Agp1 and His248 in Agp2. This seems to be a conserved motif. It might be that His260/250/248 plays a similar role in the photocycle of all three phytochromes. This is in agreement with the suggestion of the proton-transferring role of His260 in Cph1 by Velazquez et al. [27].

The findings of both Velazquez et al. [27] and Takiden et al. [28] further confirm the validity of the here presented results.

The here presented computational results make it possible to reliably simulate the CBD of Agp2 and investigate the role of His248 and His278, as well as PSC(B) and PSC(C). This will help to get further insight into dynamics inside the CBD and their possible influence on the photocycle, as will be discussed in the next chapter.

Chapter 10

QM/MM calculations and MD simulations of Agp2 methylester variant of BV

10.1. Experimental remarks

Studies of Agp2 *via* resonance Raman techniques have been carried out by the group of Hildebrandt et al.: They revealed that the thermal back reaction from Pr to Pfr is controlled by a keto-enol-equilibrium [23]. This is enabled by the additional protonation of the His278 residue at its ε-position. The origin of this proton is not quite clear, but it has been suggested that it may come from the protonated propionic side-chain (PSC) of ring C. They found, it carries a proton in the Pfr state of BV, as discussed in section 9. The corresponding signal in the RR is disappearing upon Meta-F to Pr state transition which they interpreted as possible proton translocation from PSC(C) to His278 and thus enabling the thermal back route to Pfr *via* a keto-enol-tautomerie [23]. This proton transfer highlights the importance of both PSC(C) and His278 for the correct photoconversion.

Past studies on different phytochromes revealed already a significant importance of the carboxylic groups at the cofactor for the correct photoconversion. In 2009 and 2010 Lagarias and coworkers [24],[25] presented studies of modified PSC's of the BV and PCB chromophores (DrBphP and Cph1, respectively) *via* amidation of the propionates. They investigated the impact of this on the stability of the chromophore-protein system and the correct photoconversion [24],[25]: For BV-containing *Dr*BphP the amidation of PSC(B) led to a different Pr ground state and multiple Pfr states. PSC(C) modification created dual Pr-states which could be excited to dual Pfr-states. Further, the two states seemed to collapse into one single state upon denaturation. Both differing in a denaturation-sensitive property, such as protonation. They concluded that PSC(B) seems to play an important role in sensing red light, while PSC(C) determines the width of the determined spectral region. They also tested this procedure on PCB-containing phytochrome Cph1, where amidation of PSC(B) showed a mixture of different Pr-like states, in which at least one species was able to undergo correct photoisomerization. Similar to their studies on *Dr*BphP, PSC(B) was also found to be responsible in Cph1 for defining the sensed spectral range. In contrast to this, the amidation of PSC(C) led to an inhibition of Pfr-formation. Therefore, it has to be important for the correct formation of the Pfr-adduct. Both propionates in both phytochromes have not been found essential for photoisomerization of the chromophore, but are influencing the correct Pr to Pfr transition. Especially in Cph1, PSC(C) is involved critically. They thus concluded that both PSC's are important for stability inside the protein and correct adjustment into its pocket, determination of spectral range sensed by the proteins and pKa-tuning of the surrounding amino acids [24],[25]. Formation of amides hinders this to a certain extent [24],[25].

Following these findings, Franzisco Velazquez and Maria Fernandez tested the idea of Prof. Siebert, to modify the propionic side-chains (PSC(B) and PSC(C)) of biliverdine of phytochrome Agp2 by a methylation procedure, yielding monomethylester variants of BV [26]. These esters should shed further light on the importance of the PSC's on the photocycle in general and the thermal backreaction (as described in [23]) in particular. In Figure 43 both variants are shown schematically. Here, the one with methylmonoester at ring B is called BVMB and the one with the ester function at ring C is called BVMC. The unmethylated PSC's were protonated in their studies [26]. The following experimental findings are currently unpublished [26]:

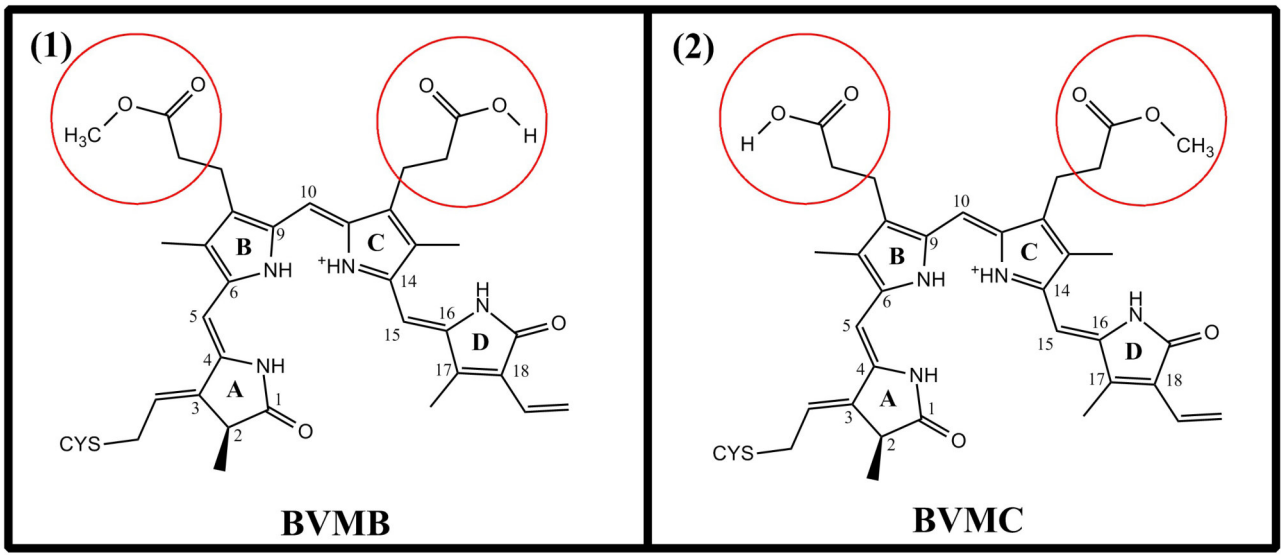

Figure 43. Variants of Agp2 cofactor biliverdine – monomethylesters: (1) BVMB: BV with methyl group at PSC of ring B and proton at PSC of ring C; (2) BVMC: BV with methyl group at PSC of ring C and proton at PSC of ring B; modified areas are encircled red.

It was not possible to get pure BVMB and BVMC probes, but only a 1:1 mixture of both which has been incorporated into the Agp2 Apoprotein. They tested the correct insertion of the modified chromophores into the Apoprottein *via* UV/VIS spectroscopy. The formation of a Pfr-like state could be observed (band at 745 nm absorption), although it formed very slowly (after 77 h it was still not completed). After irradiation with far-red light the formation of the Pr-like photoproduct could be observed in UV/VIS at 675 nm (typical absorbance of Agp2 Pr state). The thermal back conversion happened very slowly (not complete after 16 h). Also another state could be detected, appearing as a band at 685 nm. This was assigned to a photoinactive state. They concluded that two different states were formed: a photochemically active Pfr-state and one which was unable to convert into a Pr-state [26].

To further evaluate these findings, they did RR measurements of the mixture and compared the results to spectra from native Agp2 [26]: They found that the dark state (Pfr-like) showed high similarities to the native protein. They confirmed the appearance of an enol-form of BV in the Pr state upon irradiation. The RR spectrum of the Pr form was again similar to the one of the unmodified protein. Differences in RR spectra between methylated and native species could be found e.g. in the region around 1300 cm^{-1}. Here propionate-contributions are typically present. Furthermore, H/D exchange revealed a protonated chromophore at the pyrrole nitrogens, similar to the native form.

They suggested [26] that the monoesters form a mixture of 1:1 active:inactive, meaning that either BVMB or BVMC is active and the other one forms a photoinactive species. Since, they observed the Pfr to Pr thermal backconversion with a proton on PSC(C) in the corresponding RR of the Pfr, it is likely to assume that BVMB undergoes photoconversion. Also, in both Pfr- and Pr-RR-spectra, strong changes for bands including contributions from PSC(B) could be observed. This also points towards BVMB as the active species, while BVMC only forms an inactive state.

They conculded that since the methylation of PSC(C) led to an inactive form, this side-chain and its proton is essential for correct photoconversion and for the proposed keto-enol-triggered backreaction [23],[26]. In contrast to this, the methylation of PSC(B) only led to an elongated Pfr-formation period, as well as, a delayed thermal backreaction but still enables the phytochrome to

undergo correct photoconversion [26]. The phytochrome is thus more tolerable to modifications of PSC(B), while maintaining its photoactivity [26].

Theoretical studies at an atomic-level have to be carried out, to further investigate the influence of the methylation of both PSC's on the stability and structure in the CBD. Taking the findings of section 9 of the protonation states of His248 and His278 into account, it is now possible to simulate Agp2 with methylated PSC's of BV in a reliable way.

10.2. Theoretical investigations of Agp2 – BV-variants

The ed-model of Agp2 has been used in QM/MM computations and MM-MD simulations with methylated variants of BV. The models BVMB and BVMC have been built, using the crystal structure of Agp2 in its parent Pfr-state as solved by Scheerer et al. [16]. Hereby BV was either methylated at PSC(B) and protonated at PSC(C) (BVMB) or *vice versa* for BVMC. The protonation of either PSC(B) (BVMC) or PSC(C) (BVMB) was chosen to be consistent to the synthesized esters [26] as shown in Figure 43. The decision, which oxygen of the corresponding PSC was methylated or protonated, was made upon visual inspection of potential binding partners and possible space for placing the relatively large methyl group. The methyl group in BVMC has been placed on the same oxygen of PSC(C), as was the hydrogen in the computations in section 9. A third model has been built with methylation at both PSC-sites (called BiMET). With the exception of the BV-modifications, the rest of the structure in all models was completely similar. For further details please consult sections 5.2.2. and 5.2.5. In these studies the impact of different variants of BV (WT, BVMB, BVMC and BiMET) on the structural stability of the CBD has been investigated. Deviations from the crystal structure model (protonated, solvated and water(proton)-equilibrated X-Ray structure from Scheerer et al. [16]) have been analyzed. This procedure was used to validate the experimental assumption that BVMB is the photoactive form.

10.2.1. QM/MM computations of models BVMB and BVMC

Both, BVMB and BVMC models have been used in a hybrid QM/MM geometry optimization with an active region of the modified BV and surrounding protein residues and water molecules in a 20 Å sphere, centered at BV and quantum chemical treatment of BV, Arg211, His278, His248, Asp196 and 11 waters as shown in section 5.2.5. The results of these are shown in Figures 44 to 47 and Tables 9 and 10. Here, the corresponding starting structures (crystal structure with modified BV and surrounding amino acids and water molecules which have been treated quantum mechanically) are depicted in blue, while the optimized models are shown in colors. The distance-evolutions, concerning the distances between the oxygen/methyl group (carbon) of PSC(B) and the amine NH-group of Arg211 (O/CH_3(PSC(B))-NH(Arg211), as well as the distances between the oygen/methyl group (carbon) of PSC(C) and the N_ε of His278 (O/CH_3(PSC(C))-N_ε(His278)) have been also subject of investigation. Both optimized models have been studied for deviations from the original crystal structure. For this purpose the stabilities of O/CH_3(PSC(B))-NH(Arg211) and O/CH_3(PSC(C))-N_ε(His278) distances, as well as RMSD values for BV, Arg211 and His278 and His248 and the qualitative stability of the H-bond-network in 5 Å radius of the chromophore, during the optimization was investigated. All RMSD values in the following have been computed with alignment of optimized model and crystal structure at the cofactor.

In BVMB (as shown in Figure 44) the structure of BV and surrounding residues show many similarities after optimization to the starting model. The biggest deviations can be seen at the methylated PSC(B) and smaller ones at His278 (s. Figure 44). PSC(B) and the methyl group are turning to the side (to avoid sterical clash with Arg211), while Arg211 is moving away from BV.

This is also reflected in the relatively high RMSD for Arg211 of 0.88 Å (s. Table 9). An interesting feature is that although the saltbridge between the carboxyl group of PSC(B) and Arg211 is lost, an interaction between PSC(B) and Arg211 is still maintained. The movement of PSC(B) brings the non-methylated oxygen of PSC(B) nearer to the protonated amino-group of Arg211. In contrast to this, the distance between methyl-carbon and the primary amino-group of Arg211 is getting larger which can also be seen by the increase in the corresponding distance. Therefore, hydrogen-bonding between PSC(B) and Arg211 is maintained (s. Table 10), while simultaneously Arg211 moves away from BV. In contrast to this, PSC(C) shows only small deviations from the original structure, as well as His278 which is moving towards BV (decreasing distance between both) with a moderate RMSD of 0.7 Å (s. Table 9). Also here, the interaction between BV and His278 is maintained *via* H-bond (not shown here) as depicted in Table 10. Further, the positions, waters do occupy after the QM/MM optimization (not shown here) do not change from the crystallographic positions.

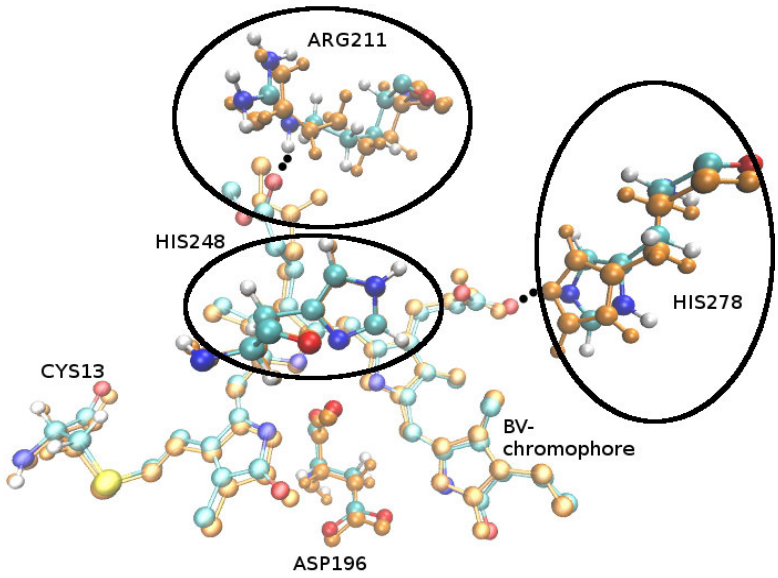

Figure 44. Optimized structure (colors) of ed-model with methylated PSC(B) compared to X-Ray [16] (orange); both aligned at chromophore (NOTE: no protons at BV are shown; **BV is protonated at PSC(C)**)

In BVMC the situation is quite different (see Figure 45): His278 moves strongly away from PSC(C) and deviates from its original crystallographic position. The reason might be the influence that the methyl group at PSC(C) has on His278. Due to the methyl group PSC(C) is less negatively charged, possibly influencing the electrostatic interaction between both, as well as it is sterically more demanding. This is also reflected in the increasing distance between both and the very high RMSD of 1.12 Å of His278 (s. Table 9). PSC(C) itself shows no movement at all. Also Arg211 is slightly shifted, as well as PSC(B). While in BVMB the introduction of the CH_3 group leads only to a rearrangement of named residues, but intermolecular interactions are maintained, it is quite different in BVMC. Here the CH_3 group on PSC(C) leads to a loss of BV-His278 connection (no hydrogen bond is directly or indirectly *via* water-bridge connecting both) and bigger instabilities in the whole structure which can be seen in the higher RSMD values (s. Table 9), as well as in the instability of the hydrogen-bond-network in BVMC (not shown here) as depicted in Table 10.

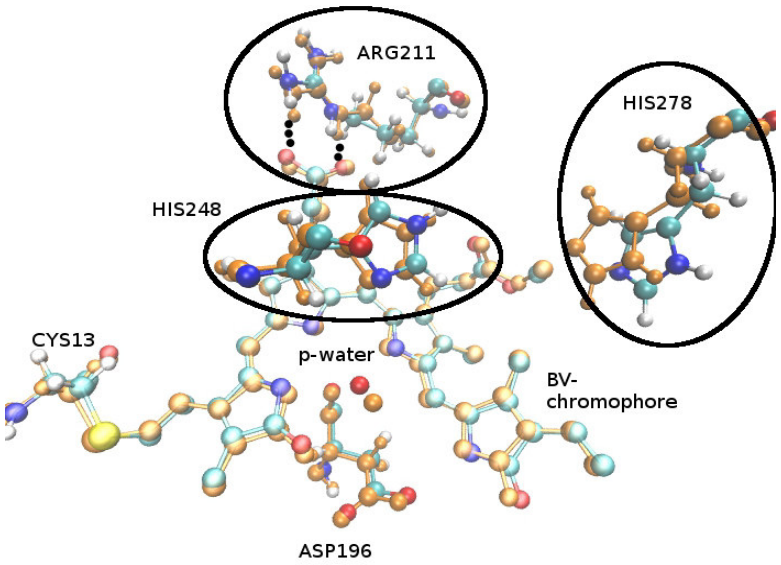

Figure 45. Optimized structure (colors) of ed-model with methylated PSC(C) compared to X-Ray [16] (orange); both aligned at chromophore (NOTE: no protons at BV are shown; **BV is protonated at PSC(B)**)

In Table 9 the computed RMSD's are shown for BVMB and BVMC (comparison to BiMET is done in the next section). As stated, BVMB shows the biggest fluctuations at Arg211 and moderate ones for all other three residues. BVMC shows similar patterns, with the exception of the high value for His278. The RMSD for Arg211 in BVMB is Δ=0.38 Å higher as in BVMC, while the corresponding RMSD of His278 in BVMC is Δ=0.42 Å higher as in BVMB. Concluding, BVMC shows overall higher deviations from the starting structure than BVMB.

RMSD (Å) (heavy atoms)	**BVMB H248E H278D**	**BVMC H248E H278D**	**BiMET H248E H278D**
BV	0.68	0.44	0.62
R211	**0.88**	0.50	0.71
H248	0.49	0.56	**0.81**
H278	0.70	**1.12**	**1.30**

Table 9. Computed RMSD values for bimethylated BV model and of both monomethylester structures BVMB and BVMC with aligned chromophore and respect to the crystal structure [16]

Additionally, the H-bond-network (s. Table 10) in BVMB remains the same during the whole QM/MM computation, while in BVMC it changes. This is especially true for the H-bond connection between PSC(C) and His278, where no H-bond can be found in the optimized BVMC model. Also no water molecules are bridging between both PSC(C) and His278. This is in strong contrast to BVMB, where a H-bond-connection is maintained between Arg211 and PSC(C).

	BVMB H248E H278D	BVMC H248E H278D	BiMET H248E H278D
Stable/not stable	✔	×	✔/×

Table 10. Comparison of H-bond-network stability in bimethylated BV model and both monomethylester structures BVMB and BVMC with respect to the crystal structure [16] (✔: stable network → no changes; ×: unstable network → many + crucial changes, concerning either R211 or H278; ✔/□ : moderate changes → e.g. only exchange of bonding partners)

This shows that methylation at PSC(C) leads to higher instabilities in the chromophore binding-pocket and loss of BV-His278 interaction which might induce an photoinactivity of BVMC as postulated by Hildebrandt et al. [26] The interaction of the protonated PSC(C) and His278 thus might be essential for correct photoconversion (as suggested in RR experiments [23],[26]). BVMB on the other hand shows moderate deviations from the crystal structure and maintained the original H-bond network, even between Arg211 and BV. Thus, Arg211 is more flexible and tolerant towards the artificial introduction of the methyl group and initiates an opening-up of the CBD, in contrast to His278 for methylation at PSC(C). Additionally, His248 shows almost no deviations from its starting positions ans is seemingly not affected by the single methylation at either PSC.

10.2.2. QM/MM computations of model BiMET with bimethylated BV

These findings have been validated by similar computations of a bimethylated species of BV. QM/MM geometry optimizations with an active region of the modified BV and surrounding protein residues and water molecules (in a 20 Å sphere, centered at BV with QM-treatment of Arg211, His278, His248, Asp196, BV-variant and several water molecules; see section 5.1.3. for more details) have been carried out and the results have been analyzed in terms of structural deviations (as compared to X-ray structure [16]) and compared to the findings for BVMB and BVMC. The calculations yielded the structure depicted in Figure 46.

No proton translocations can be observed during the optimization procedure. Furthermore, Arg211 and His278 deviate from their original positions in the crystal structure. The former shows moderate deviations similar to the model BVMB. It moves slightly away from BV. His278 shifts strongly away from the cofactor as already shown in BVMC. The distance-evolutions (not shown here) show that both Arg211 and His278 are moving away from the cofactor, resulting in larger distances between these residues and PSC(B) and PSC(C), respectively. In contrast to this His248 shows almost no deviations from its original crystal position (according to Figure 46). In contrast to this it gives an RMSD value of 0.81 Å (s. Table 9) which is much higher than the ones in BVMB (about 0.32 Å) and BVMC (about 0.25 Å), respectively. The reason for this might be the contradicting movements/directions of Arg211 and His278, as will be discussed in more detail in the next paragraph, as well as in section 10.2.4. Arg211 moves in another direction as His278, and since both are connected via beta-sheet (in which also His248 is located) their movements are pulling at His248 in different directions, resulting in this high RMSD without seeing big deviations in Figure 46. In contrast to this, the chromophore structure is mostly maintained. Only PSC(B) shows some rotations, due to a rearrangement with Arg211 to circumvent steric contacts between methyl group and Arg211. This movement is exactly the same as observed for BVMB. Again the hydrogen bonding interaction between Arg211 and PSC(B) is maintained (s. Table 10). In contrast, no H-bond could be observed between BV and His278 (s. Table 10). This residue lost completely the contact to the cofactor which may be the reason for its strong movement, since no attractive H-bond-interactions between both are maintained. Furthermore, the crystal waters in the vicinity of the cofactor are mostly undisturbed (not shown here).

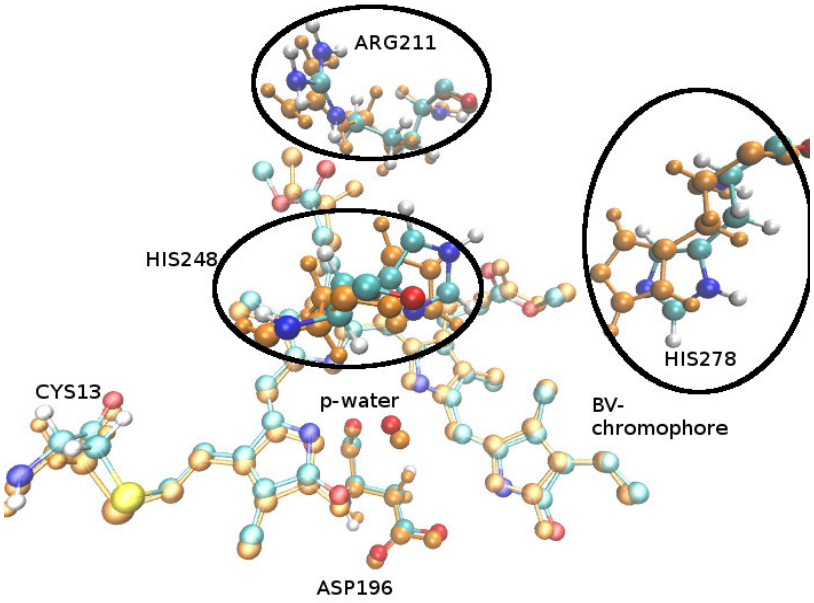

Figure 46. Optimized structure (colors) of ed-model with bimethylated BV at PSC(B) and PSC(C) compared to X-Ray [16] (orange); both aligned at chromophore (NOTE: no protons at BV are shown)

These observations reflect a mixture of movements seen in BVMB and BVMC models. Arg211 is moving upwards and away from BV (with respect to BV, as shown in Figure 46), just like in BVMB. But the movement is interrupted by the translation of His278. This is triggered by the CH_3 group at PSC(C) and is similar to BVMC. This induces a shift of His278 which counters the opening-up of the chromophore binding-pocket, as triggered by the Arg211 translations. Therefore, trends of both BVMB and BVMC are at least partly visible in BiMET, but neither is fully established, since both are opposing each other. In Table 9 the corresponding RMSD values in comparison to previous ones for BVMB and BVMC are depicted. Arg211 has a higher RMSD of 0.71 Å in BiMET than in BVMC, but lower than in BVMB. This shows again that Arg211 may be hindered by movements of His278. The RMSD of the latter is even higher (1.30 Å) than in BVMC, indicating even bigger deviations and fluctuations than in the monomethylester. The reason for this may be that on one hand the translocation of Arg211 triggers the shift of the connected beta-sheet β1 (of which His278 is a part of; for further analysis of secondary structure elements inside the CBD see sections 1.2., Figure 3 and 10.2.4. and 10.2.5.). This induces a directed movement of His278. On the other hand, the methyl group on PSC(C) induces a translation of His278 into another direction, away from BV. Both movements are increasing the distortions of His278 which leads to an increased RMSD of His278 even higher than in BVMC. Also His248 shows stronger fluctuations (as reflected by an RMSD of 0.81 Å) than in the mono-variants. This is (as pointed out in the paragraph before) a direct consequence of the two controversial movements induced by Arg211 and His278 which are initiated by the introduction of the two methyl groups at the PSC's. Furthermore, the qualitative stabilities of the hydrogen-bonding-networks are shown in Table 10 for BiMET and compared to BVMB and BVMC. In the bimethylated variant changes in the hydrogen bonds occur always near the methylation sites. In more detail, the interaction between Arg211 and PSC(B) is ultimately maintained, while His278 loses every contact to BV (even no water is bridging between them). Again this reflects a mixture of the findings in the mono esters.

Conclusively, BiMET shows a mixture of trends observed in BVMB and BVMC. Both RMSD's and H-bond-network, as well as visual inspection of the optimized structure, are showing an upward

movement of Arg211 and at the same time a shift of His278 away from BV. Both movements may be hindering each other. This results partly in very high RMSD's for His278 and His248. Again, Arg211 demonstrate its ability to compensate the disturbance of the CH_3 group and still interacts with BV. This is in contrast to His278 which is less tolerant to the modification of PSC(C). These results support the findings and conclusions of the previous section 10.2.1.

10.2.3. MM-MD of Agp2 WT (with ed-protonation of His248 and His278)

The motions of Agp2 WT protein with unmodified BV has been simulated with a MD-simulation: the crystal structure with ed-protonation of His248 and His278, as well as a proton on PSC(C) has been used for the MM-MD-computations (with a production run of 20 ns; all-atom-free, with the exception of BV, for which all heavy atoms have been fixed; for further details of used methods, please see section 5.2.1.). The comparison to the variants BVMB and BVMC will be discussed in more detail in section 10.2.5.

RGYR and RMSF calculations of protein (not shown here in detail)

The radius of gyration changes during the production run from values around 29 Å to smaller ones around 28 Å. This might reflect folding of parts of the protein (mainly PHY region). RMSD computations show an equilibrated protein structure at the end of 20 ns production. According to RMSF calculations the highest flexibility lies within the PHY domain with values up to 6 Å. Arg211 and His278 have both values of approximately 1 Å.

Structure of CBD

Of high interest are the movements and stabilities of the two residues Arg211 and His278. As has been shown in the QM/MM computations (see sections 10.2.1. and 10.2.2.), the methylation of PSC(B) and PSC(C), respectively, has direct impact on these two residues. Therefore, the CBD in general and these two residues especially, have been analyzed more thoroughly. The RMSD's of both residues stay constant at ~ 1.5 Å (His278) and ~ 0.5 Å (Arg211) with the exception of small fluctuations (during the simulation). After 20 ns of simulation time, His278 is at a distance of 4 Å to the hydrogen of PSC(C). Arg211 shows a distance of 2 Å to the oxygen of PSC(B).

As can be seen in Figure 47 (left) Arg211 does stay very near its initial crystal position at the end of simulation. His278 on the other hand shows strong deviations from its original position. It moves away from the chromophore during the production run and further the imidazole ring is turning around 180°. These movements are in contrast to the observed behavior of His278 in the corresponding QM/MM computations as discussed in section 9.2., where His278 showed almost no shift at all from its initial crystal position. The reason for this will be discussed at the end of this subsection. The motions (and possible deviations from the X-ray) of protein residues His248, Arg242, Tyr205, Pro459 and Arg456 have also been studied in more detail. As depicted in Figure 47 (left), His248 shows a deviation in its orientation towards BV. Arg242, which is a direct interaction partner of Arg211 (as can be seen in Figure 47 on the left), stays stable at its initial position, only following the occasional movements of Arg211 by adjusting the corresponding H-bonds between them. Tyr205 only reorientates, turning its OH-group away from BV in the process, but staying in the same distance as in the crystal structure (not shown here for better visualization). For Arg456 and Pro459 the situation is different (Figure 47 on the right):. Both residues are shifted away from BV. Arg456 shifts towards Pro459 and the latter moves in the same direction. Their initial orientation towards BV remains the same at the end of the simulation. Therefore, for both the distances to their original neighbors (Arg456: CH_3 of ring A of BV; Pro459: C=O of ring D of BV)

at BV are increasing in a significant way (a possible reason for this will be explained in the next paragraph). All these movements occur in the beginning of the MD and afterward the residues stay stable.

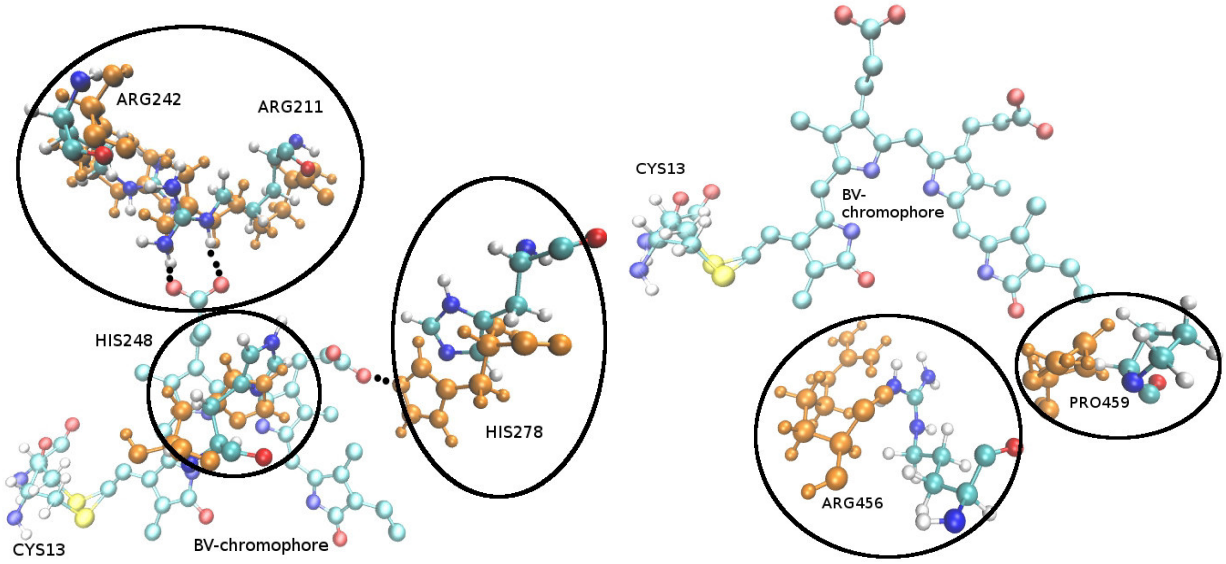

Figure 47. BV and surrounding protein residues: (left) Arg211, Arg242 and His278, His248; (right) Arg456 and Pro459; after 20 ns MD-simulation (colors) of WT variant of BV compared to X-Ray [16] (orange); both aligned at chromophore; all-atom free, except fixed heavy atoms of BV (NOTE: no protons at BV are shown; **BV is protonated at PSC(C)**)

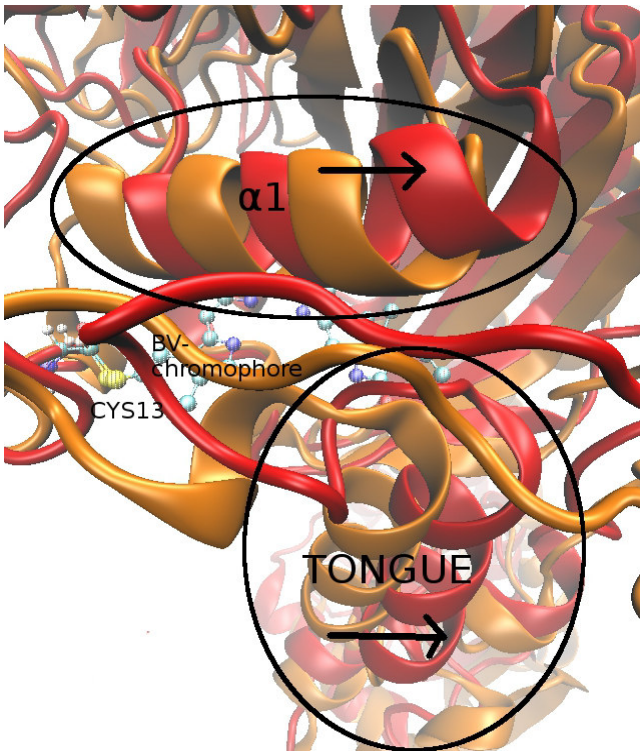

Figure 48. Comparison of secondary structure elements between crystal structure [16] (orange) and structure after 20 ns of MD-simulation (red) of CBD; arrows indicate direction of shift of secondary structure elements occurring during MD-simulation of WT; all-atom free, except fixed heavy atoms of BV; aligned at BV (NOTE: **BV is protonated at PSC(C)**)

Changes in secondary structure of the CBD are depicted in Figure 48. The alpha helix α1 (located directly above BV; in red) is shifted away from its original position (in blue). Furthermore, as visualized in Figure 48, the tongue, as well as beta-sheets deviate in the same direction. This is in accordance to the observed movements of Arg456 and Pro459, since they are located in the tongue-region. These structural elements are shifted to the right (in respect to the position of BV in Figure 48, left). These deviations occur already in the beginning of the MD, while for the rest of the simulation these residues stay stable at their new positions. Therefore, the residues seemingly reached their maximum deviation from their original positions (which will play an important role when investigating BVMC with MM-MD, see sections 10.2.4. and 10.2.5.). The PHY domain is more compact after the MD (as the movements in the CBD trigger the rearrangement of the PHY domain). This confirms the findings of RGYR computations which reflected folding of the globular protein.

In summary, Arg211 does not deviate form the crystal structure, while His278 and His248 show displacements and shifts away from the cofactor. The CBD displays an opening-movement, as α1, β1, β2 and the tongue are translated away from their original crystal positions (and away from BV), into the same direction (to the right, with respect to BV, as depicted in Figure 48). A reason of this strong deviation can be the conditions of the MD-simulation. All heavy atoms of BV have been remained fixed during the simulation which might cause a rearrangement of the surrounding protein residues. BV is not able to adjust to the CBD and this might be the origin of the observed movements in the CBD. Further that might be why the results from the MM-MD are differing from the results in the respective QM/MM computations in section 9.2. The observed movements are limited in the way that after a certain distance to their respective initial positions is reached, the secondary structure elements remain stable on their new positions. The rest of the CBD shows only small deviations. Although these results are not comparable to the QM/MM combinations (due to fixed BV), they can and will be used as a reference for comparison with BVMB- and BVMC-MM-MD-trajectories (since here, also all heavy atoms of BV have been fixed during the MD-simulation). Thus the influence of the introduced methyl groups on the respective PSC's on the here discussed motions and the general stability of the CBD can be observed.

10.2.4. MM-MD of models BVMC and BVMB and comparison to WT

MM-MD simulations have been carried out for Agp2 photosensory domain with variants of BV. The chromophore has been either methylated at PSC(B) or PSC(C), named BVMB and BVMC, respectively (as already discussed e.g. in section 10.2.1.). For comparison with the findings of the QM/MM computations and the MM-MD-simulation of the WT (sections 10.2.2. and 10.2.3., respectively), 20 ns production runs have been done (all-atom free, with the exception of fixed heavy atoms of BV; for further details on e.g. heating, equilibration and simulation conditions, please see section 5.1.1. and 5.1.2.) for both models and compared to each other. In the following, the results obtained by these simulations are shown and compared to each other and to the findings for MM-MD calculations of Agp2 WT (s. section 10.2.3.).

RMSD and RMSF calculations for MM-MD's

The radius of gyration of BVMB is increasing over the course of the MD, from 28.5 Å to 29.5 Å which reflects an opening movement of the protein (which is in contrast to the folding, observed in the WT simulation). In contrast to this the RGYR in BVMC is decreasing from 29 Å to 28 Å, indicating an increased folding of the protein (similar to the WT). The RMSD of BVMB is stable after 5 ns of simulation time, while the RMSD of BVMC is not stable after 20 ns. The latter model is not completely equilibrated (as shown in the next paragraph). RMSF values for BVMC (protein)

are higher than in the WT (up to 9 Å for the PHY domain), while BVMB shows smaller ones (up to 4.5 Å in the PHY domain). These observations display a higher flexibility and instability inside the CBD of BVMC (as found also in the QM/MM calculations; s. section 10.2.1.), probably induced by the methylation of PSC(C).

Structure of CBD

Since Arg211 and His278 are in near vicinity of the artificial methyl group, both have been subject to more detailed investigations. Analysis of RMSD's for both residues in the two models BVMB and BVMC revealed for the former stable values around 0.5 Å for Arg211 and 0.2 Å for His278 at the end of the simulation (with only minor fluctuations during the MD for His278 which settle in the end of the MD). The situation is quite different for BVMC. Here both residues display higher RMSD's of ~1 Å for Arg211 and 1.2 Å for His278 at the end of the MD. These values are not stable, but change continuously over the coarse of the simulation. A similar trend is displayed by the distance-evolutions between both protein residues and their respective PSC's (PSC(B) for Arg211 and PSC(C) for His278): In BVMB, after initial increasing in both distances, both remain stable till the end of the simulation. In the end the distance between CH_3 on PSC(B) and NH (side-chain) of Arg211 is ~6 Å, while the one between H on PSC(C) and N_ε of His278 is ~4 Å. This indicates a strong rearrangement of Arg211 and a smaller one for His278. In BVMC both distances fluctuate during the whole simulation around 4 Å. Especially His278 switches between two distinct conformations (differing in the orientation of the imidazole ring towards BV) which might reflect heterogeneity of His278. In Figure 49 the positions of both Arg211 and His278 at the end of the MD are depicted in comparison to the original crystal structures for both BVMB and BVMC.

Arg211 is moving away from BV in BVMB (Figure 49, left), while maintaining its original orientation towards BV. With 6 Å it is too far away for formation of H-bonds with the cofactor, but this huge distance might be attributed to the fixation of all heavy atoms of BV which might otherwise adjust to the movement of Arg211. His278 shows also a small displacement from the X-ray structure, with remaining orientation towards BV. In BVMC (Figure 49, right), Arg211 shows almost no shift from its original position (the fluctuations in RMSD and distance are attributed to internal motions, like rotations around single-bonds). His278 on the other is strongly deviated from it starting position. It moved strongly away from BV and also the imidazole ring is rotated about 90° (it is continuously rotating during the MD and occupying two different conformations differing of about 90°). As a first conclusion, Arg211 is more flexible and can better adjust to the CH_3 group at PSC(B) and remains stable on its new position. His278 on the other side is not able to adjust to the methylation of PSC(C), remaining unstable and changing between two different conformations.

Interestingly in the BVMB model, His248 is moving also away from BV (following the motion of Arg211), showing a strong deviation from its crystal position (see Figure 49, left). Since in BVMC His248 does not show such stark movements (s. Figure 49, right), it is has to be a result of the methyl group at PSC(B). Other residues which are affected by the introduction of the methyl group are Tyr251 (stationed between both PSC's) and Arg242 (directly interacting with Arg211). Arg242 is pictured in Figures 49 for both models. In BVMB Arg242 is moving strongly and following the translation of Arg211, away from BV, while maintaining the hydrogen-bonds with Arg211. At the end of the MD it stays stable. Tyr251 shows also a displacement movement in the same direction as Arg211 in BVMB and a rotation of ~90° around the central axes of its ring in BVMC. The MD-simulation of BVMC is different in the way that here R242 shows less deviations from the initial structure. Nevertheless it fluctuates strongly during the MD-simulation, thus staying unstable till the end. The same is true for Tyr251 remaining unstable.

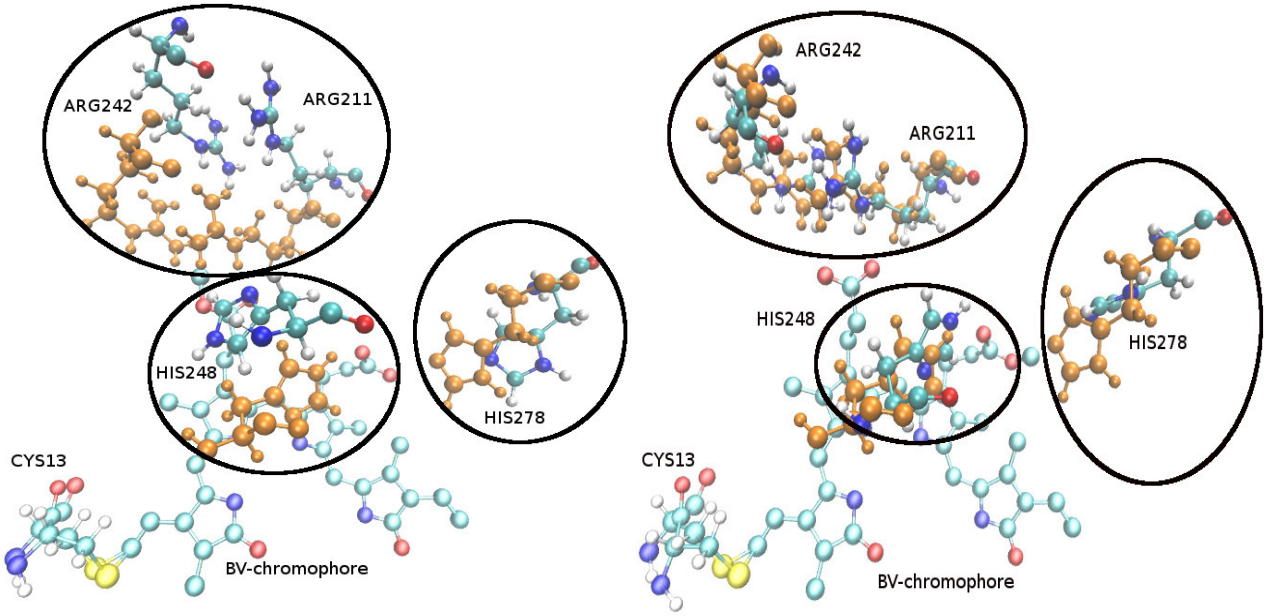

Figure 49. BV and surrounding protein residues Arg211, Arg242 and His278 after 20 ns of production run (colors) (all-atom free, except fixed heavy atoms of BV; aligned at BV) of ed-model compared to X-Ray [16] (orange): (left) BVMB (**BV is protonated at PSC(C)**); (right) BVMC (**BV is protonated at PSC(B)**); no protons are shown at BV

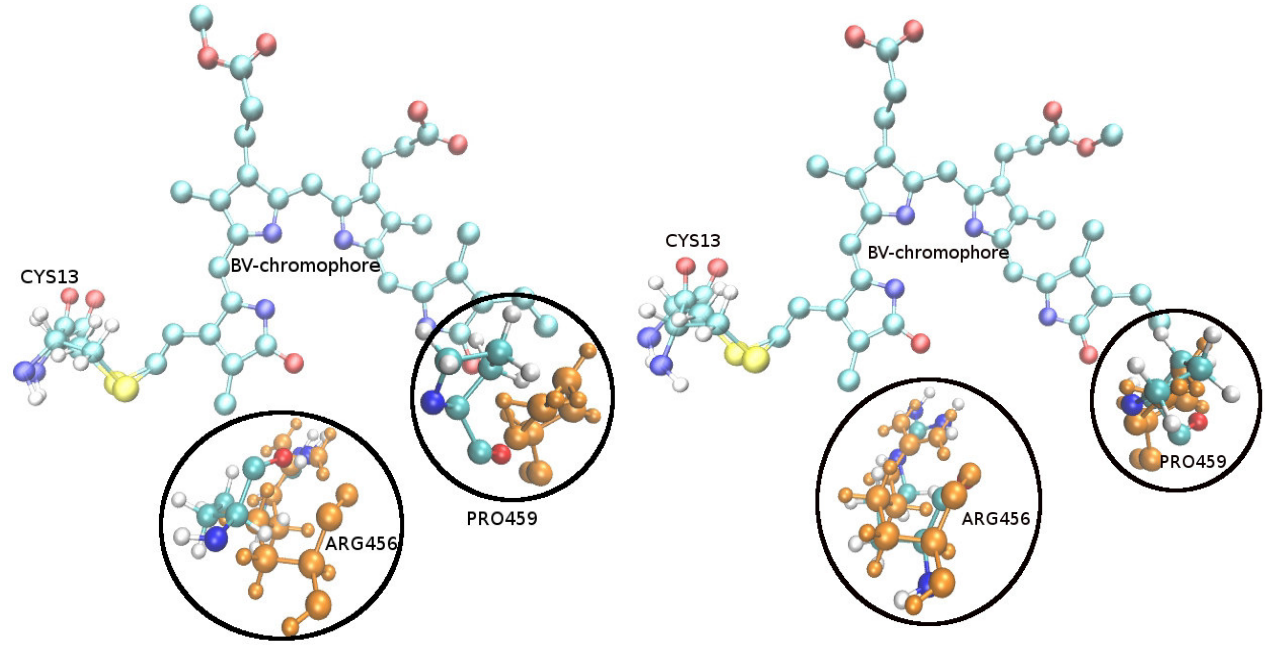

Figure 50. BV and surrounding protein residues Arg456 and PRO459 after 20 ns of production run (colors) (all-atom free, except fixed heavy atoms of BV; aligned at BV) of ed-model compared to X-Ray [16] (orange): (left) BVMB (**BV is protonated at PSC(C)**); (right) BVMC (**BV is protonated at PSC(B)**); no protons are shown at BV

Both Arg456 and Pro459 (depicted in Figure 50 for BVMB (left) and BVMC (right)) show no sign of translation away from its original position and remain unchanged. This is in contrast to the displacement of both residues in the WT (see section 10.2.3.). The methylation of both PSC's seems to reduce the degrees of freedom of protein residues Arg456 and Pro459 or to enhance the

intermolecular interactions between them and BV.

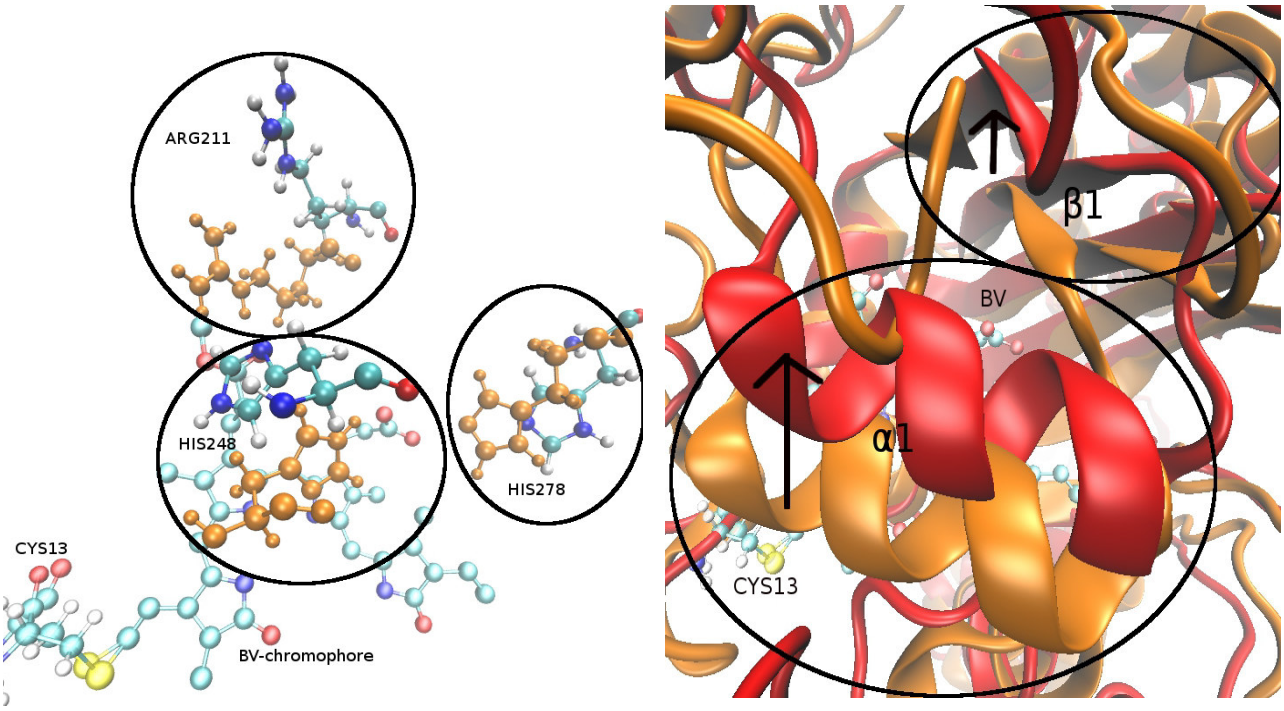

Figure 51. Structural comparison between crystal structure [16] and structural motifs after 20 ns of MD-simulation of BVMB (all-atom free, except fixed heavy atoms of BV; aligned at BV): (left) BV, His248, His278, Arg211 (X-ray: orange; BVMB: colors); (right) CBD with BV and comparison of secondary structure elements between X-ray structure [16] and BVMB at the end of the simulation (X-ray: orange; BVMB: red), arrows indicate direction of shift of secondary structure elements occurring during MD-simulation in methylated variant; (NOTE: no protons are shown at BV; **BV is protonated at PSC(C)**)

Looking at the secondary structure, BVMB reveals a directed movement of protein residues located in the CBD (as shown in Figure 51): α1, β1 and β2 are displaced away from BV and their original position (upwards, with respect to BV and Figure 51, right). The reason for this is the adjustment of Arg211 to the CH_3 group on PSC(B). Arg211 is located in β1. The CH_3 group on PSC(B) induces the translation of Arg211 which forces the complete β1 strand to follow its movement. His278 is also located in β1 explaining its deviation from the crystal structure. β1 is further connected to α1 (located above BV, as seen in Figure 51, right) following the translation of β1. Summarizing, the area around modified BV is opening like the lid on a barrel, thus it will be called barrel-like opening from here on. In this way the CBD adjusts to the introduced CH_3 at PSC(B). Therefore, Arg211 has a high flexibility to adjust to changes at the PSC(B). These changes in the CBD lead to an equilibrated and stable (meaning, without bigger fluctuations of the protein residues) CBD structure. This observations could already be made in the corresponding QM/MM computations and are in agreement with these (see section 10.2.1.).

Such a barrel-opening cannot be seen in BVMC (see Figure 52). Here the biggest deviation observed, is the one for His278 which is shifted to the right. All investigated residues showed continuous fluctuations during the MD-simulation. These movements do not settle at the end of the simulation (in contrast to BVMB). Secondary structure element analysis in near vicinity of the cofactor show no obvious deviations from the initial structure (see Figure 52, right). The exception is a slight displacement of α1 in near vicinity of His278.

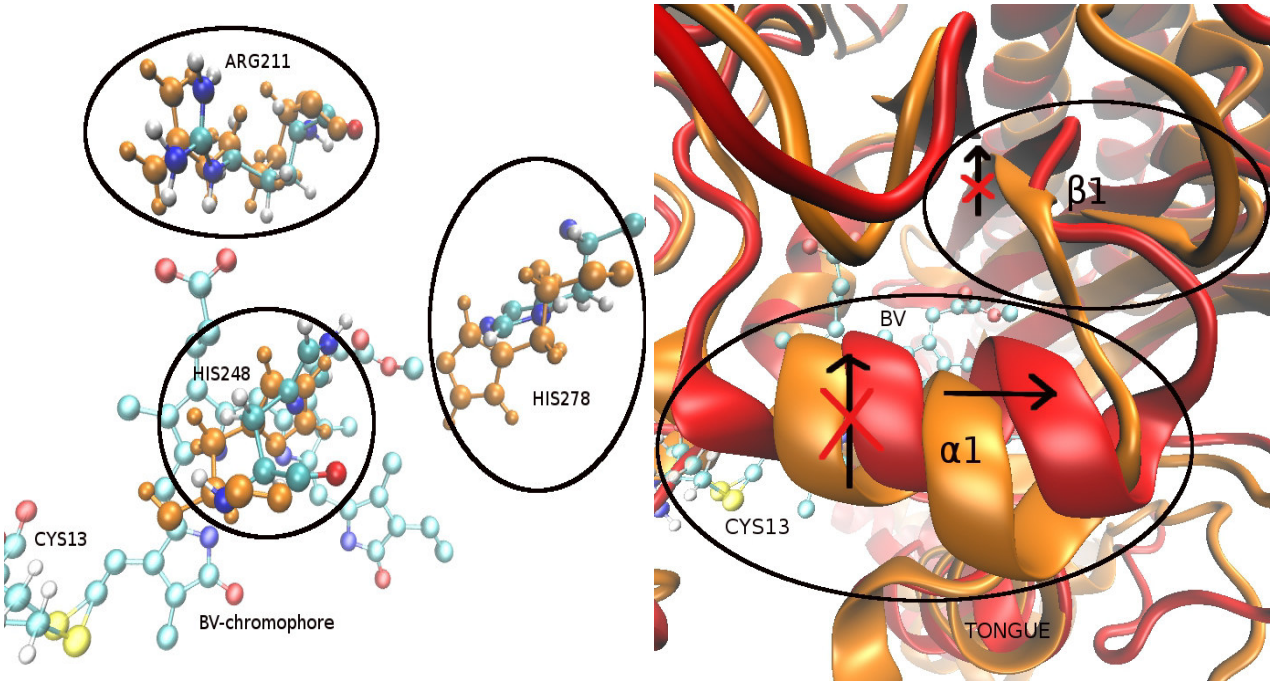

Figure 52. Structural comparison between crystal structure [16] and structural motifs after 20 ns of MD-simulation of BVMC (all-atom free, except fixed heavy atoms of BV; aligned at BV): (left) BV, His248, His278, Arg211 (X-ray: orange; BVMC: colors); (right) CBD with BV and comparison of secondary structure elements between X-ray structure [16] and BVMC at the end of the simulation (X-ray: orange; BVMC: red), arrows indicate direction of shift of secondary structure elements occurring during MD-simulation in methylated variant; (NOTE: no protons are shown at BV; **BV is protonated at PSC(B)**)

In summary, BVMC shows high and lasting instabilities, as though the structure is not able to adjust to the CH_3 group at PSC(C). Subsequent QM/MM geometry optimizations of the last frame of the MM-MD-simulations for both models with free BV showed no changes at all of the structures which means that the here discussed structures at the end of the MD-simulations represent structures in an energy well or are at least close to an energy minimum. Therefore, these QM/MM results of the end-structures of the trajectory are not shown and discussed here.

MD-simulations of BVMB/BVMC VS. MM-MD of WT

The MD-simulations of BVMB and BVMC have been compared to the results for the MD of the WT in terms of structural changes in the global structure, as well as in the direct vicinity of BV. In Figures 53 and 54, the structures (aligned at BV) of BVMB (colors/turquoise) vs. WT (blue) or BVMC (colors/turquoise) vs. WT (blue) at the end of the MM-MD-simulations are compared to each other.

In Figure 53 (left) it is depicted that the WT is showing a strong displacement for residues Arg456 and Pro459 away from BV and their crystallographic positions, compared to BVMB. His278 is in both WT and BVMB shifted away from BV. In the BVMB model, residues Arg456 and Pro459 are staying at their initial position. The reason for this could be the strong movement of Arg211 and α1, β1 and β2 in BVMB which lead to steric hindrance for Arg456 and Pro459 to follow the more native movement as observed in the WT. Further, Arg211, His248 and Tyr251 are moving away

from BV in a direction almost perpendicular to the one of Arg456 in the WT (upwards, with respect to BV and as depicted in Figure 53 on the left). In Figure 53 (right) it is shown that in BVMB secondary structure motifs α1, β1 and β2 are displaced to different positions, as compared to the WT. While BVMB shows a barrel-like opening of the CBD, neither can be observed in the WT-simulation and WT-structure at the end of the simulation.

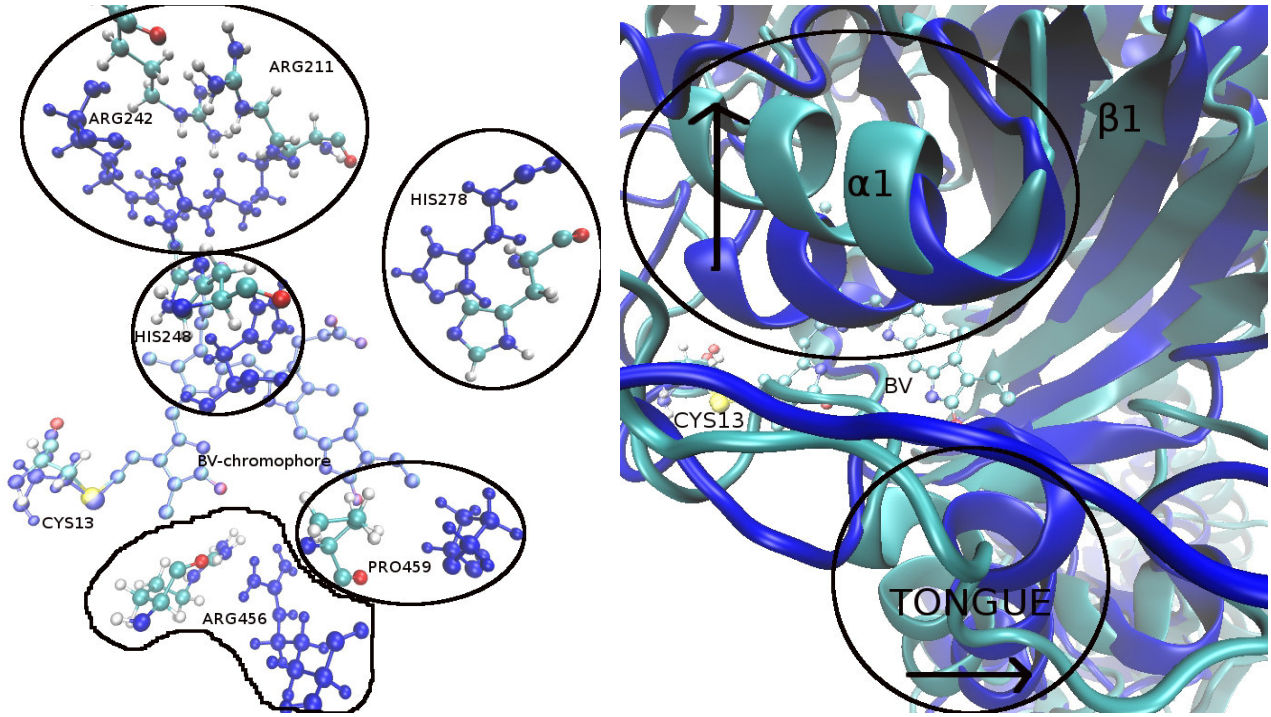

Figure 53. Comparison between structural motifs after 20 ns of MD production runs for BVMB and for WT (all-atom free, except fixed heavy atoms of BV; aligned at BV): (left) BV + Arg211, Arg242, Arg456, Pro459, His248 and His278 (BVMB: colors; WT: blue); (right) CBD with BV and secondary structure elements (BVMB: turquoise; WT: blue), arrows indicate direction of shift of secondary structure elements occurring during respective MD-simulations; (NOTE: no protons are shown at BV; **BV is protonated at PSC(C)**)

In the case of BVMC the situation is different: In Figure 54 (left and right) many similarities can be seen between both WT and BVMC after MD-simulation. For example Arg211, His248 and Tyr251 are nearly at the same position in both models. The difference between both WT and BVMC is the displacement of residues Arg456 and Pro459 in the WT. This is not visible in BVMC. A reason for this could be increased steric hindrance at these positions, due to the CH_3 group at the PSC of ring C. In both MDs, His278 is shifted away from BV in the same direction, but shows a rotational variation in BVMC, in contrast to the WT.

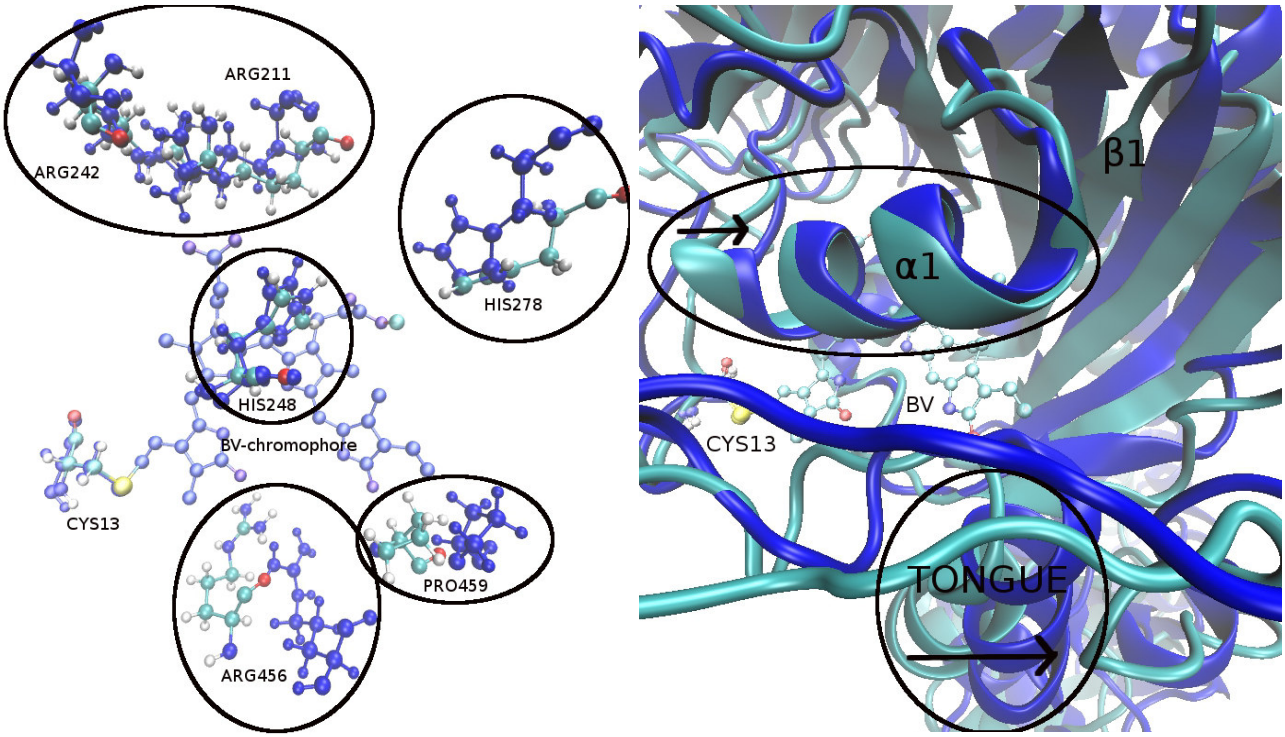

Figure 54. Figure 53. Comparison between structural motifs after 20 ns of MD production runs for BVMC and for WT (all-atom free, except fixed heavy atoms of BV; aligned at BV): (left) BV + Arg211, Arg242, Arg456, Pro459, His248 and His278 (BVMC: colors; WT: blue); (right) CBD with BV and secondary structure elements (BVMC: turquoise; WT: blue), arrows indicate direction of shift of secondary structure elements occurring during respective MD-simulations; (NOTE: no protons are shown at BV; BV is protonated at PSC(C) in WT and at PSC(B) in BVMC)

10.2.5. Discussion of MM-MD-simulations of BVMB and BVMC models

In summary, the methylation of PSC(C) leads to a very unstable BV binding-pocket. Especially His278 is shifting continuously. Already in the trajectory of the WT His278 is moving away from BV (since in all three discussed MD-simulations, the heavy atoms of BV have been kept fixed, they are comparable to each other). This movement is limited to a certain distance between His278 and BV, after which the shift is stopped. After introduction of CH_3 on PSC(C) His278 seems to be "uncomfortable" near to it which is why it is displaced from its original position. But, for finding a stable position it might need to go even farther away from BV as in the WT which it cannot. It has no more room to go. Therefore, it continuously changes between two conformations. This instability might be furthered by the missing hydrogen bonding between BV and His278 (direct and indirect *via* water) and possible electronic repulsion between CH_3 and His278 (CH_3 is more negatively charged than the native hydrogen at this position). This indicates a low tolerance of His278 towards changes at the PSC(C).

BVMB on the other hand shows adjustment of the surrounding protein residues of BV to the methylation of PSC(B). The barrel-like opening movement, triggered by the rearrangement of Arg211 leads to an altered but stable structure of the CBD around BV.

Although BV has been kept fixed in the discussed MD-simulations of BVMB and BVMC and the WT, the discussed results for the two former ones are in agreement with the observations made for the all-atom free QM/MM optimizations, as discussed in section 10.2.1. Since, in the simulation of

the WT the fixation of BV induces deviations in the CBD from the crystal structure, which are in contrast to the all-atom free QM/MM optimizations, the deviations of the CBD during the simulation of BVMB and BVMC have to be induced by the introduction of the methyl groups at PSC'S. These alterations of the chromophore structure dominate the structural rearrangements of the corresponding CBD's during the trajectory of both methylated models. Methylation induced motions dominate the ones, induced by the fixation of BV.

Conclusively it may be assumed that BVMB might be the photoactive species (as suggested [26]), with Arg211 triggering the adjustment to the CH_3 group on PSC(B) and the subsequent stabilization of the CBD. In contrast to this, BVMC shows possible heterogeneity of His278 and high instabilities due to the CH_3 group on PSC(C) in the CBD. Further, the connection (H-bond) between His278 and PSC(C) is lost. Therefore, it may be photoinactive, as proposed by experiments [26].

The significance of these findings are first, the importance of the His278-PSC(C) interaction for correct photoconversion. It is very sensitive and not tolerant to modifications. Second, the high tolerance of Arg211 for adjustments in its connection with BV is in strong contrast to the His278-BV interaction. Arg211 can accept modifications int the CBD and still maintain the necessary intermolecular interaction with the cofactor. Additionally, His248 may be important for the thermal back reaction, since in the probably active species BVMB its position and orientation towards BV is strongly altered which may led to the slow back conversion observed in the experiments [26]. For further conclusions and possible future work please consult chapter 11.

10.2.6. MM-MD of model BiMET with bimethylated species

The model with bimethylated biliverdine BiMET has also been subject to a 20 ns long MD-simulation (all-atom-free, except fixed heavy atoms on BV; for further details of model-building, as well as simulation conditions please see section 5.2.3.) to get an impression of the structural changes inside the BV binding-region. The influence of the simultaneously attached methyl groups at both propionates of BV on the stability of the CBD and possible residue-displacements are investigated and the obtained structure at the end of the simulation is compared to the crystal structure [16].

RGYR and RMSF calculations for MM-MD's

The RGYR is decreasing during the MD-simulation from ~28.8 Å to 28.4 Å which reflects a folding of parts of the protein (like observed in BVMC). The constant RMSD of the protein during the simulation indicates an equilibrated structure at the end of 20 ns production run (with similar values ~3.5 Å as obtained for BVMB). RMSF calculations yielded very high values for the CBD (up to 5.5 Å) indicating strong movements in this area and which are higher than in BVMB. Nevertheless, the highest values (observed in the PHY domain) of about 6.5 Å are in between the ones for BVMB (4.5 Å) and BVMC (8.5 Å), maybe due to a combination of trends of both models which will be discussed in the following.

Structure of CBD

Protein residues Arg211 and His278 are investigated more closely, since they are in direct connection to the corresponding PSC's and their methyl functions. Interestingly, for Arg211 very high RMSD values (up to ~2 Å) have been observed (higher than in both monomethylester models. It moves away from BV and stays stable at a distance of ~6 Å (similar to BVMB) between NH of Arg211 and CH_3 on PSC(B). The high RMSD values are due to a further shift in orientation of

Arg211. Arg211 (NH-group of the side-chain) turns away from Arg242 (depicted in Figure 55 on the left). Arg242 (side-chain) is also turning away from Arg211 and does not follow its motions, as has been found in BVMB and BVMC. Although there is still hydrogen bonding between both arginine residues (between C=O of Arg242 and the NH of the side-chain of Arg211), it may be not sufficient enough to stabilize Arg211 on its new position which may be the reason for its high RMSD values.

His278, as well as His248, show almost identical behavior as observed in the simulation of BVMC. His278 is displaced away from BV in the same direction as in BVMC (translation occurred already in the thermalization phase, thereafter it stays stable). Additionally, a rotation around the central axes of the imidazole ring like in BVMC could be observed (see Figure 55, left). At the end of the simulation both His278 and His248 show a similar distance and orientation towards BV as found for BVMC. His248 is rotated and shifted away from BV, in comparison to the crystal structure. Additionally, Tyr251 is also shifted away from their crystal positions and stay at almost the same positions, as observed for BVMC. None of these residues show any kind of movement as found in BVMB. Arg456 and Pro459 stay at their respective starting positions, like it have been seen for both mono-variants.

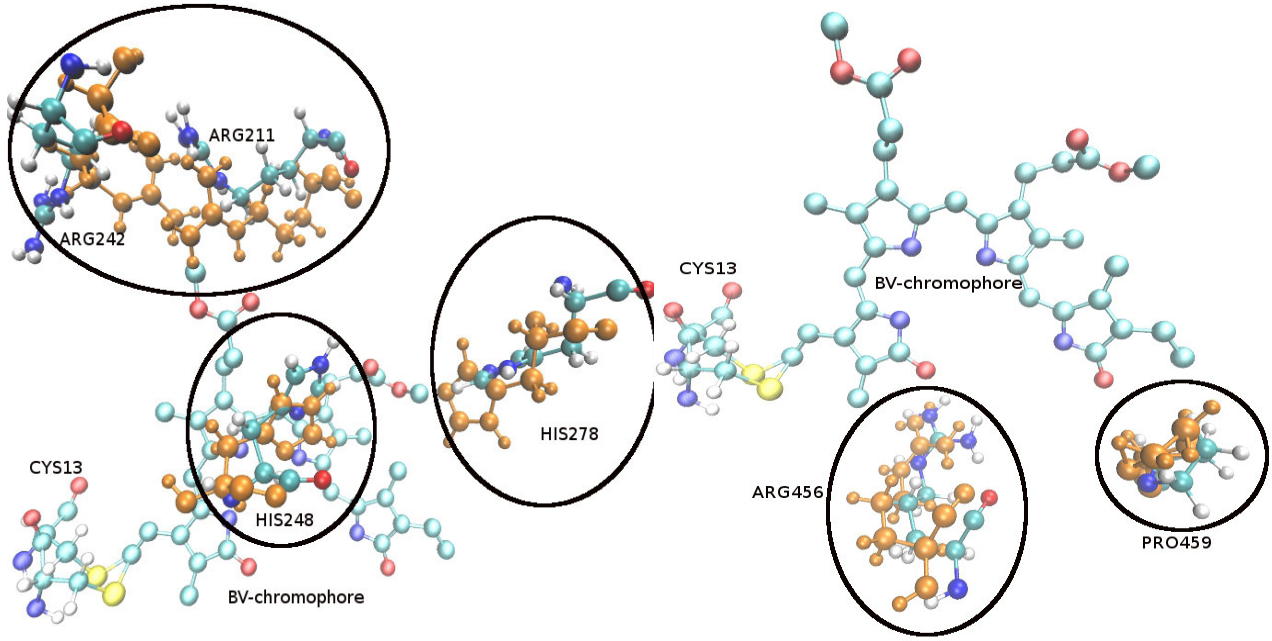

Figure 55. BV and surrounding protein residues of BiMET: (left) Arg211, Arg242 and His278; (right) Arg456 and Pro459; after 20 ns MD-simulation (colors) of bimethylated variant of BV compared to X-Ray [16] (orange); both aligned at chromophore; all-atom free, except fixed heavy atoms of BV (NOTE: no protons at BV are shown; **BV is methylated at both PSC's**)

As a first conclusion, the movement of Arg211 seems to be disturbed by the translations of protein residues induced by the CH_3 group on PSC(C) (as already discussed for the corresponding QM/MM optimization of the bimethylated species; see section 10.2.2.). On one hand, this induces higher instabilities on Arg211, but on the other hand the movement of Arg211 also influences the stability of His278 in a positive way. This can be explained by looking at the secondary structure elements inside the CBD as depicted in Figure 56 on the right side.

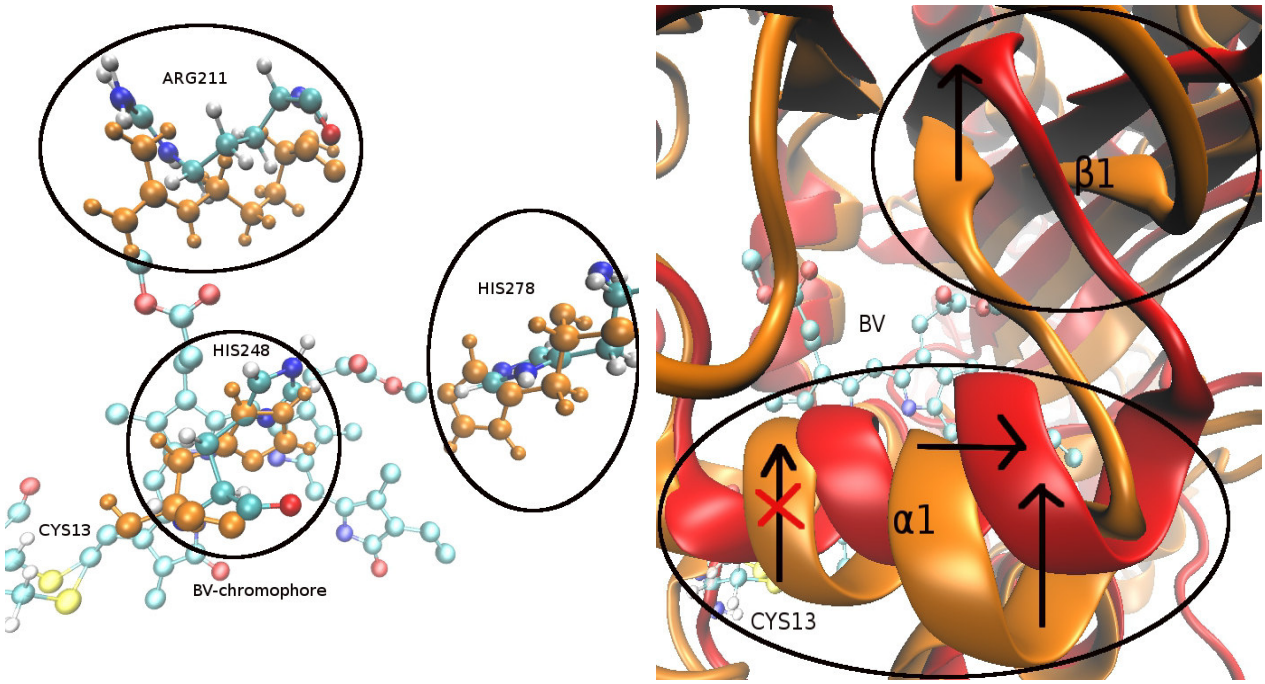

Figure 56. Structural comparison between crystal structure [16] and structural motifs after 20 ns of MD-simulation of BiMET (all-atom free, except fixed heavy atoms of BV; aligned at BV): (left) BV, His248, His278, Arg211 (X-ray: orange; BiMET: colors); (right) CBD with BV and comparison of secondary structure elements between X-ray structure [16] and BiMET at the end of the simulation (X-ray: orange; BiMET: red), arrows indicate direction of shift of secondary structure elements occurring during MD-simulation in methylated variant; (NOTE: no protons are shown at BV; **BV is methylated at both PSC's**)

As visualized in Figure 56 (right), both α1 and β1 are showing displacements similar to the ones observed for BVMB (upwards, with respect to BV, as depicted in Figure 56, right). But, additionally, to this barrel-like opening of the pocket (triggered by the translation of Arg211) a second movement in a direction perpendicular to it (to the right of BV, as depicted in Figure 56, right; like in BVMC) is occurring. These movements are interfering with each other. The opening of the pocket triggered by PSC(B) and Arg211 is hindered by the methyl group at PSC(C) and the induced movement of His278. The CBD still opens up more as in BVMC, but less, as observed in BVMB. Therefore, the barrel-opening of the CBD region is interrupted. In contrast to this His278 is more stable on its position (with no visible heterogeneity) which may be due to the influence of Arg211 and the motion of β1 on the movements of His278. Additionally, it is important to mention that subsequent QM/MM calculations of the structure at the last frame of the MM-MD of BiMET yielded no changes at all, thus the structure at the end of the MD-simulation has to be inside an energetic well or near to an energy minimum which is why the corresponding results are not shown here.

10.2.7. Discussion of MM-MD-simulations of BiMET model

Since all investigated structural parameters for BiMET showed similarities to either BVMB and BVMC (as discussed also in the QM/MM section 10.2.2.) it can be concluded that the incorporation of two methyl esters at the propionate side chains of BV lead to a mixture of BVMB and BVMC induced movements which influence each other. The interference of these with each other lead to higher strains on Arg211 and the displacement of Arg242. Overall it yields a final structure in between the ones for BVMB and BVMC at the end of their respective MD-simulations: Arg211 and

PSC(B) behave similar as in BVMB, while His278 and PSC(C) behave similar to their counterparts in BVMC. Thus, the results of MM-MD-simulation and QM/MM calculations for the BiMET model, validate the trends observed for the monoesters, due to the partly reproduction of the trends of the monomethyl-variants.

Chapter 11

Summary and conclusions of part IV

Protonation states of amino acids are important for structural stability and dynamics and even mechanistic properties of proteins. The assignment of these to titratable amino acids is not trivial and may be influenced by the presence of prosthetic groups. Since X-ray analysis is not able to predict protonation states of amino acids, it is important to find other ways for their correct prediction. Especially histidines are a major challenge, because they can obtain several possible protonation patterns which experiments, like vibrational spectroscopy or theory might provide. In the case of Agp2, a bathy phytochrome, for which its crystal structure has recently been solved by Scheerer et al. [16], the protonation state of two histidines in near vicinity of the cofactor, His248 and His278, remained unclear. These two histidines might have a very important role in the photocycle. Therefore, the correct assignment of their protons has been of great importance for further studies of the photocycle of Agp2. Hybrid QM/MM approaches have been used for analysis of nine different protonation models (differing in the position of the proton on either N_ε or N_δ or on both nitrogens of both His) based on the crystal structure of Agp2 in its Pfr state [16].

The model with the highest stability (lowest RMSD-values and least deviations in the H-bond-network) and which yielded an QM/MM-optimized structure with the least deviations from the initial X-ray structure was the one with His248 carrying a proton on N_ε and His278 on N_δ, yielding the ed-model. All other models showed higher deviations from the original crystal structure and are thus less probable to represent the native protonation states of His248 and His278 in the WT of Agp2. Further, it could be shown that the protonation patterns strongly affect the stability of the chromophore and its surrounding residues. The protonation of His278 at its N_ε always resulted in structures with strong deviations of His278 from its crystallographic position and also complete loss of its connection to BV *via* H-bonds.

Furthermore, the computations of models with deprotonated PSC(C) resulted in the model with His248 protonated at N_ε and His278 doubly protonated as the one with the least deviations from the crystal structure. A strong tendency of a proton transfer from N_ε of His278 to the oxygen of the PSC of ring C could be observed. Such a proton translocation would lead back to the ed-model with protonated PSC(C). This further underlines the findings concerning the ed-model being the most probable native protonation state of His248 and His278 in the Pfr-state of Agp2. Furthermore, this demonstrates the validity of recent resonance Raman findings [23] which propose that the PSC(C) carries a proton.

Therefore, QM/MM geometry optimizations and the analyzed structural parameters (like RMSD's and H-bond network stability), are valuable tools for obtaining protonation motifs, like for the discussed histidines and may be of help in future investigations of similar problems.

These findings have been used for the building of the three different models BVMB, BVMC and BiMET, in which the propionic side-chains of BV have been methylated either on PSC(B) (BVMB), PSC(C) (BVMC) or on both (BiMET). QM/MM and MM-MD computations of these models and subsequent comparison to the crystal structure, as well as the native structure with unaltered BV revealed more details on the properties of BV-surrounding protein residues Arg211 and His278 and both PSC's and their possible influence on the stability of the CBD. The results are shown in Table 11.

Model	Experimental RR-Results [26]	Calculation Results
BVMB	• Pfr-Pr-transition observed with H^+ on PSC(C) • bands of PSC(B) undergo big changes in RR of Pfr- and Pr-states (1288 + 1305 cm^{-1}) → CH_3 on PSC(B) → **proposed photoactive species** ✔ • very slow formation of Pfr-like state (745 nm absorbance) • irradiation: Pr-like state (675 nm absorbance) → slow back-conversion	• Arg211 moves away from BV • maintains H-bond to PSC(B) • triggers movement of α1 and β1 away from BV → barrel-like opening of CBD • equilibrated • Arg211: high tolerance for CH_3 → least deviations from X-ray [16] → **proposed photoactive species** ✔
BVMC	• photoinactive state (685 nm absorbance) → no Pr-state formation observed → **proposed photoinactive species** ✘	• higher RMSD's and RMSF's as BVMB • His278 switches between two different conform. + strong deviations from X-ray-position [16] • not equilibrated • H-bond between His278 + PSC(C) is not maintained → **proposed photoinactive species** ✘
BiMET	----------	• RMSD's and RMSF in between BVMB and BVMC • Arg211 less flexible • His278 deviates from X-ray-position [16] • no H-bond between His278 + PSC(C) • water-bridge in between Arg211 and PSC(B) → barrel-like opening is hindered → **mixture of BVMB + BVMC**

Table 11. Summary of experimental [16],[26] and computational results for BV-variants BVMB, BVMC and BiMET in comparison to each other – determination of possible photoactive species BVMB and inactivity of BVMC

The different computations of the three models and their comparison revealed the influence of the different methlyations on the structural integrity of the CBD. The two residues Arg211 and His278 have been of higher interest, since they are in direct contact to the two propionates. It was found that Arg211 is able to compensate modifications at PSC(B): It moves away from BV (in BVMB) and induces a barrel-like opening of the BV-surrounding pocket by translation of alpha-helix α1 and beta-sheet β1 which stay stable at the end of the corresponding MD-simulation. Further QM/MM

calculations revealed the maintenance of the hydrogen-bond between Arg211 and PSC(B) of BV. Also the connection between the proton on PSC(C) and N_ε of His278 is maintained. In summary BVMB shows the least deviations from the crystal structure, as opposed by the BVMC model: Here, His278 was not able to adjust to the methylation at PSC(C). This model revealed strong deviations for His278 from its initial crystallographic position. His278 moves away from BV, but is not able to adjust to the CH_3 group. That leads to His278 continuously changing between two different orientations and high RMSDs. Also, RMSF-values for all protein residues in near vicinity of BV are drastically higher in BVMC than in BVMB. Finally, also the direct interaction between His278 and PSC(C) is lost (since no H-bond is connecting both) which is in contrast to the BVMB model. The computations of the bimethylated species BiMET confirms these findings, since it represents a mixture of trends observed in both monoesters.

Therefore, it can be concluded that methylation of PSC(B) induces a barrel-like opening of the CBD, indicating the tolerance of Arg211 to disturbances, while methylation of PSC(C) leads to increased strains and deviations of His278 from its crystal position. Also BVMB shows an equilibrated structure after 20 ns of MM-MD-simulation, while the same is not true for BVMC. Thus, incorporation of BV with CH_3 on PSC(B) leads to a more stable CBD, with the least distortions from the X-ray structure [16].

Furthermore, these findings lead to the assumption/proposal that BVMB may be photoactive (due to the ability of Arg211 to adjust to a modified PSC(B) and the stable CBD after described barrel-like opening) and BVMC is photoinactive (due to loss of PSC(C)-His278 connection and instabilities in CBD). This is in agreement with experimental results [26], in which results from RR-studies suggested a complete photoconversion for BVMB, while BVMC has been proposed to be unable to undergo the Pfr-Pr-transition correctly (s. section 10.1.). Further these results underline the importance of the PSC(C)-His278 interaction and the presence of the proton on the PSC(C) for stabilization of the cofactor inside the CBD and correct photoconversion. Additionally, the experimentally observed [26] relatively slow thermal back-conversion from Pr to Pfr for the photoactive species can be explained with the BVMB model: Here a strong displacement of His248 (as observed during the MD-simulation of BVMB) may be the reason for this slow dark-reversion, since possible proton-transfers (as discussed e.g. by Mroginski et al. [27]) between both HIS248 and BV will be more challenging and potentially more time-consuming.

This should be further studied by future investigations of the influence of methylated BV-variants on the structure of the CBD in the Pr-state (e.g. for Agp1) both experimentally and theoretically, with QM/MM- and MM-MD-methods. Furthermore, calculations of Raman-spectra of both Agp2 Pfr-state and Agp1 Pr-state and comparison to experimental RR-spectra would shed further light into the matter. Finally, since the MM-MD-calculations have been carried out with fixed heavy atoms of the chromophore (because of insufficient parametrization of BV in its Pfr-conformation) future investigations should also target MM-MD-simulations with free BV. Therefore, new parameters for BV in its Pfr-state should be developed and implemented into the calculations.

Conclusively, the usage of QM/MM- and MM-MD-methods revealed new structural features in the CBD of Agp2. The protonation state of His248 and His278 could be verified as ed-protonation, as well as the importance of the His278, Arg211 and the two propionates PSC(B) and PSC(C) for the photocycle. Arg211 has been revealed as highly tolerant to structural disturbances at the BV-site, in contrast to His278. These results may help future studies to understand the dynamics in the CBD and the photoconversion of Agp2 in more detail.

Part V.

QM calculations and subsequent Raman-spectra computations *via* NMA for *Rc*FDH

Chapter 12

General remarks on investigations of FDHs + Refinement of the cofactor structure

12.1. General remarks

The catalytic activity of *Rc*FDH and Mo-driven FDHs in general is important for industrial usage, such as the generation of solid fuels out of gaseous carbon dioxide. Understanding the mechanistic properties of this class of enzymes is important. As stated in the introductory part, all known FDHs harbor a molybdenum-containing cofactor (molybdenum-containing bisguaninedinucletotide-pterin cofactor, short: Moco) at the active site, at which the oxidation process is occurring. The metal atom is coordinated by two dithiolene moieties which are further connected to an pterin ligand and a guaninediphosphate (GDP) unit, each. A cysteine (Cys) or selenocysteine (Se-Cys) is possibly directly ligated to molybdenum in its +VI oxidation state (MoVI) and, furthermore, an unknown sixth ligand which can be an O*-, OH-, S*- or SH-group. Molybdenum exist in three different oxidation states (+VI, +IV and +V). Dithiolene and the pterin moieties are assumed to play a role in changing the electronic field around the metal and influence its redox potential [51].

This section and its subsections focus on preparatory operations for FDH-investigations. In detail, it summarizes already available information on the structure and mechanism of Mo-containing cofactor in FDHs for the oxidation of formate. This will display the open questions regarding the functionality of FDHs which will be targeted in this work (s. also section 3.3.). Further, this section explains the investigative approach used in this work combining experimental vibrational analysis with computational methods. Finally preparing work regarding cofactor-size and method-choice will be discussed here.

12.1.1. State of art – available results in the literature and remaining questions

12.1.1.1. State of art

Crystal structures are only available for a very small number of corresponding enzymes such as the FDH of *E. Coli* [52],[56],[182],[183]. Here MoVI coordinates besides the two dithiolene moieties a hydroxyl-group [56] or sulfido-/thiol-group [52] (depending on the study) and a selenocysteine, while MoIV is pentacoordinated without ligation of the hydroxy-group to MoIV [56] or without the selenocysteine [52]. The interpretations of these structures are unfortunately prone to errors: For example Raajmakers et al. [52] had to reinterpret the crystal structure of *E. coli* FDH from Boyington et al. [56], due to a wrong resolution of a protein loop region in near vicinity of the molybden-containing cofactor in the latter. This led to a new assignement [52] of a sulfido-(or thiol-)ligation to both MoVI and MoIV, instead of the hydroxy-group, as proposed earlier [56]. Furthermore, they showed the dislocation of Cys from the metal-center upon metal-reduction, yielding a pentacoordinated MoIV-state [52].

Kinetic assays further revealed that FDH's, like *Rc*FDH [48] and *Dd*FDH [54], are capable of both oxidation of $HCOO^-$ and reduction of CO_2.

In contrast to the X-ray-studies [52],[56],[182],[183], theoretical calculations by Mota et al. [53]

suggested, an inactive hexacoordinated MoVI state, in which both a selenocysteine and a sulfido-group are ligated to the molybdenum. They proposed a mechanism, in which Se-Cys is leaving its direct ligation-site of MoVI *via* temporary formation of a disulfide bridge with the sulfido-ligand present at the metal-site [53]. This leads to a catalytic species, in which an active form of the cofactor, in which MoVI is only pentacoordinated, with two dithiolene moieties and one sulfido-ligand [53]. The details of which will be discussed in chapter 13 of this work.

Furthermore, EXFAS analysis of *Rc*FDH [50] suggested two different hexacoordinated MoVI-states which differ in the possible ligation of either a sulfido- or an oxo-ligand and of which the later one is not catalytically active anymore [49],[50]. Also a bound Cys could be seen in both MoVI-states of the cofactor, while they were suggesting that Cys leaves the direct coordination sphere upon formate approaching the metal and reduction of MoVI to MoIV [50]. They further proposed that formate directly binds to the now free coordination-site at the reduced metal [49]. Furthermore, the two genes FdsC and FdhD where identified to be important for both sulfuration of the cofactor and its insertion and stabilization in the active center [48],[50]: In their absence the enzyme shows no catalytic activity. It has been suggested that especially FdsC is important for exchanging the oxo-ligand at MoVI by a sulfido-group [50]. This would then activate the enzyme. These findings led to the conclusion that the sulfido-ligand is important for the oxidative procedure [48].

Furthermore, EPR analysis [54],[73],[196],[197] as well as XAS-studies [50] suggested the intermediate formation of a MoV-species and showed the presence of a proton in the second coordination sphere. The possible presence of a SH-ligand at MoV has been discussed [50],[54]. This led to the assumption that the sulfido-ligand may be responsible for a hydride abstraction from formate [54].

Generally accepted in the recent publications [49],[54],[55],[72] is that a sulfido-ligand and a (Se-)Cys-ligand are directly coordinated to the metal center of the molybdenum-containing cofactor in its oxidized +VI form. In contrast to this, it is still debated, whether the MoIV-form is only pentacoordinated, with sulfido-ligation and missing Cys-ligand [49],[55] or hexacoordinated, with bound cysteine at MoIV and possible formation of a thiol ligand [54],[72].

Here, in this chapter the possible structures and different coordination spheres of the molybdenum in the MoVI- and MoIV-states are investigated. Controversies in possible sulfido-, thiol-, oxo-or hydroxo-ligation to MoVI and MoIV, as found in recent publications, are discussed in a new light.

The different mechanisms, concerning the role of the individual ligands of molybdenum, as well as the cysteine- and formate-ligation to Mo and the involvement of a MoV-species are discussed in detail in chapter 13.

12.1.1.2. Remaining questions regarding structure of Mo-containing cofactor in FDHs and mechanism of FDHs

Although it is presently assumed, further evidence will be needed to verify, if in the MoVI-form a sulfido-ligand is present, as well as in the MoIV-form. Similarly, the presence of the Cys as direct coordination partner for MoVI and MoIV, as well as the possible direct ligation of formate to the metal during the reaction cycle are controversially discussed in the literature. Finally, the structure of the inactive desulfo form, which is present in the absence of the FdsC gene [49],[50], has not been clarified.

This work will focus on the structure (chapter 12) and mechanistic properties (chapter 13) of the

molybdenum-containing cofactor, the so-called Moco. The questions targeted in this work are what groups are ligated to the molybdenum in MoVI, MoV (chapter 13) and MoIV: Whether a sulfido-ligand is present in both oxidized (both active and inactive forms, section 12.2.) and reduced states (sections 12.3. and 13.4) of the Moco or not. Is Cys ligated to the metal in the oxidized (section 12.2.) and reduced Moco (sections 12.3. and 13.3.1.)? Furthermore, changes in the first ligation sphere which might occur during the catalytic cycle (such as on- and off-going of Cys) will be investigated in this work (sections 13.2. and 13.3.2.). The question, which part of the Moco or surrounding protein-residues is responsible for the abstraction of a hydrogen from formate will also be targeted (section 13.3.1.). And is formate directly ligated to the metal, and if so, in which oxidation state of the metal (section 13.5.)? Finally, a step by step electron-dislocation via a MoV-species will be discussed (section 13.4.). The answers to these questions will lead to the generation of a new mechanistic understanding of the catalytic oxidation of formate at the Moco-site (sections 13.6. and 13.7.).

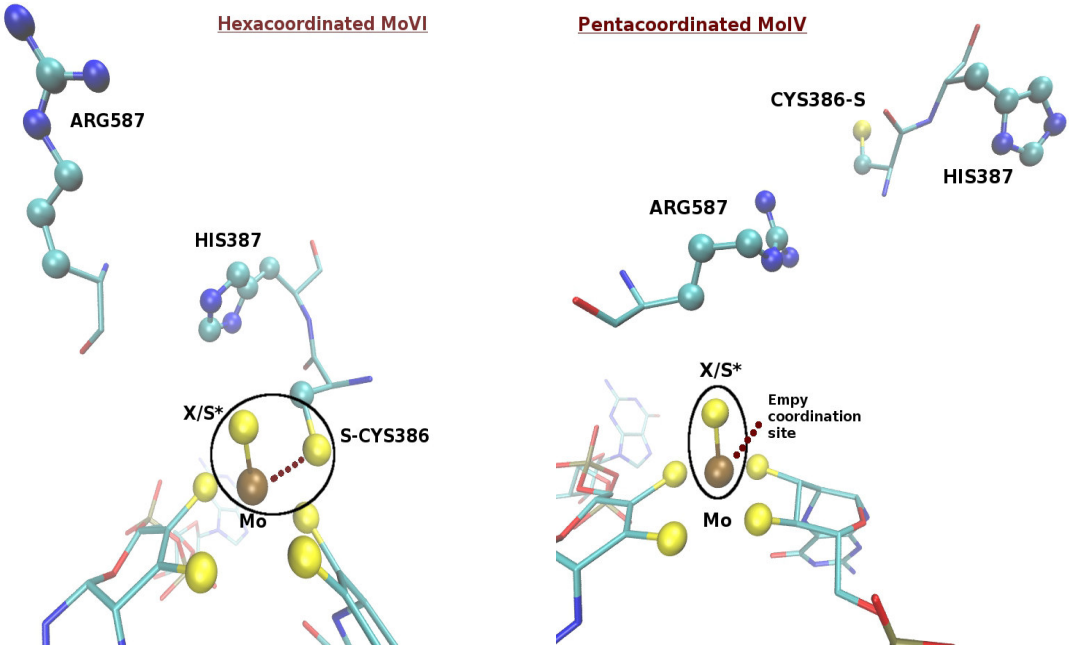

Figure 57. Recently discussed parts of the ligation-sphere of the Molybdenum in Mo-cofactor-containing FDHs (regarding S*- and Cys-ligation of Mo) for (left) oxidized *Rc*FDH and (right) reduced *Rc*FDH; shown are the structures of the homology model of *Rc*FDH by Hartmann et al. [49]; the conserved residues Arg587 and His387 are displayed: X/S* denotes either a S, SH, O or OH-ligand (in *Rc*FDH it is assumed to be inorganic sulfur S*) (pictures inspired by Fig.1 in [53])

Determination of the correct geometric state, as well as mechanistic properties of the Moco might improve the understanding of the catalytic functionality of FDHs and might even be transferable to other Moco-containing proteins with similar functions. Therefore, in the following a combination of both Resonance-Raman (RR)-spectroscopy and DFT-based QM calculations has been used to study the structure of the Moco and its possible mechanistic properties in order to answer these questions. For this purpose several RR-spectra have been measured [73]: All experimental RR/Infrared(IR)-data as well as Figures 58-60 have been thankfully provided by Stefan Wahlefeld of the group of Prof. Peter Hildebrandt [73]. For completion: Proteinsamples were provided by and biochemical experiments on RcFDH (like, e.g. protein purification and analyses [48]) have been carried out in the group of Prof. Silke Leimkühler [73]. EPR-measurements on RcFDH have been done by Benjamin Dufus [73],[196]. Model-complexes for the Moco (not shown here) have been synthesized by Ricardo Nowack in the group of Carola Schulzke [73].

12.1.2. Resonance Raman experiments of *Rc*FDH

The RR-spectra have been taken under different conditions for both the oxidized (MoVI) and the reduced (MoIV, possibly also MoV) species as is pointed out in the respective figures [73]. Concerning the naming of vibrational modes and to distinguish between the Mo-S vibration of metal-dithiolene interactions and the ones due to the sulfido-ligand, the latter will be denoted by an asterix (S*; the same for oxo-ligation O* at the metal-site).

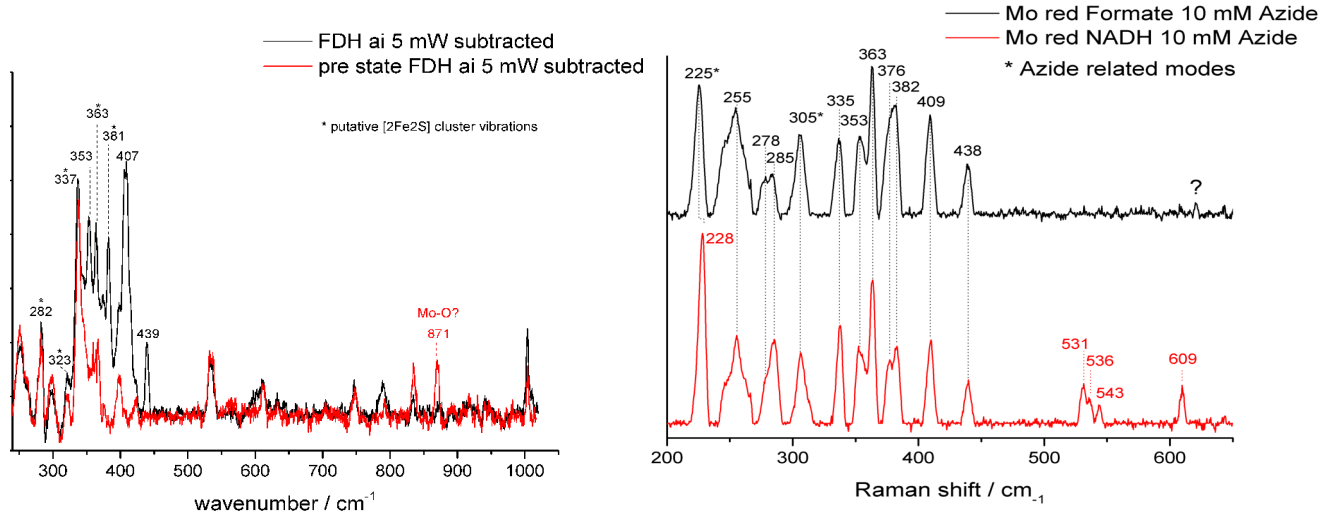

Figure 58. Experimental RR-spectra (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide) of (left) *Rc*FDH oxidized WT (black) and desulfo form (red) (subtraction spectra; both measured at 5 mW); (right) *Rc*FDH reduced form as isolated (ai): black: reduced with 10 mM formate, red: reduced with 10 mM NADH (RR-data and pictures provided by Stefan Wahlefeld) [73]

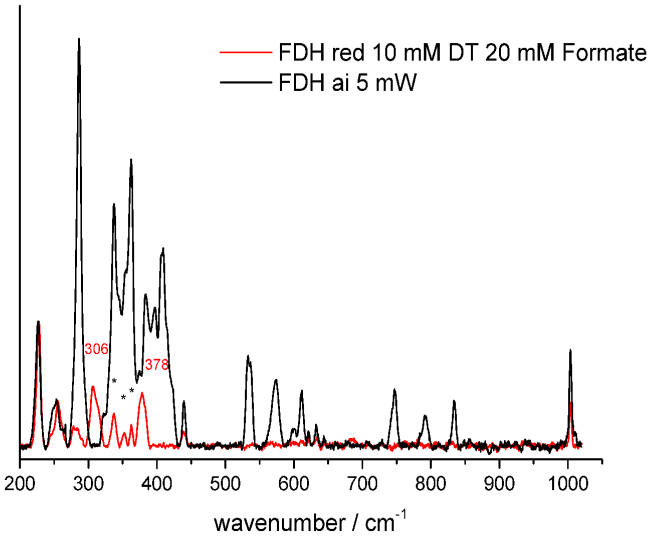

Figure 59. Experimental RR-spectra of *Rc*FDH (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT) reduced form (red) and *Rc*FDH WT form ai (black) (RR-data and picture provided by Stefan Wahlefeld) [73]

In Figure 58 on the left side, two subtraction spectra are shown for both MoVI-forms (active wildtype (WT) and inactive desulfo form (prestate)) [73]: The RR-spectra of the Apo-protein (protein without any cofactor present) have been subtracted from the as isolated ones (ai) to display only cofactor bands. The bands at 282 cm^{-1}, 323 cm^{-1}, 337 cm^{-1}, 363 cm^{-1} and 381 cm^{-1} are most

probable iron-sulfur-cluster vibrations (FeS). The two spectra which represent the active wild type form (WT) and the inactive desulfo form, differ mostly in two bands: Only in the WT-spectrum the band at 439 cm^{-1} can be seen (which may be attributed to a sulfido-metal (Mo-S*) stretching vibration, since these occur in this region), while only in the other spectrum a band appears at 871 cm^{-1} (which can be a Mo-O* stretching). The suggestion has been that the only difference between both species lies within the sixth ligand X at the MoVI (see sections 12.2.1 and 12.2.2).

On the right side of Figure 58 RR-spectra for the reduced MoIV ai form of the enzyme are shown, measured under different conditions (either reduced by formate or NADH; both in the presence of azide as stabilizing ion) [73]: In the bottom spectrum the bands after 500 cm^{-1} are most probably due to the NADH-reducing agent. The most important band in both is the one at 438/439 cm^{-1} which could represent a sulfido-ligand at the MoIV. Interestingly, the observed band at 439 cm^{-1} is at the exact position, as the corresponding one in the RR-data of the MoVI WT (see also Figure 59 [73]), maybe also indicating leftover presence of oxidized enzyme in the probe measured for the MoIV RR-spectra. But for the purpose of this study (and as will shown in the next section also proven by theoretical studies), the band has been assumed to arise from the reduced enzyme only. Furthermore, the band at ~405 cm^{-1} in all reduced spectra was assumed to arise from metal-ligand vibrations inside the Moco as well, as will be discussed in sections 13.3.1. and 13.4. It should also be mentioned that in all spectra (regardless of oxidation state) the bands before 378 cm^{-1} could probably arise from FeS-cluster-, as well as dithiolene-vibrations.

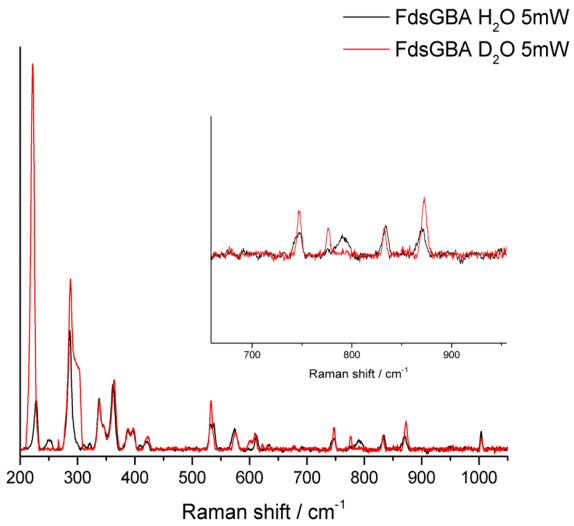

Figure 60. Experimental RR-spectra obtained for H/D exchange in desulfo-form of oxidized *Rc*FDH: red: in D_2O, black: in H_2O (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide (RR-data and picture provided by Stefan Wahlefeld) [73]

Additionally, H/D exchange experiments [73] have been carried out for oxidized (desulfo-, as well as WT-form; former shown in Figure 60), as well for reduced enzyme (not shown here): The H/D-experiments for the WT of the oxidized enzyme and the reduced *Rc*FDH yielded no shifts of bands at 439 cm^{-1} [73]. From this, it has been concluded that it might be very improbable that a SH-ligation in both oxidized WT and reduced form of the Moco can be found (further comparison with computations can be found in sections 12. and 13., as well as in the Appendix). Since a small shift of the band ~400 cm^{-1} in the H/D-RR-spectrum of reduced species could be observed, the contribution of a deuteration-sensitive species in the reduced probe (possibly only shortliving) cannot be completely ruled out. This will be discussed in further detail in section 12.3. Regarding the Moco in the desulfo form of the protein, which has been suggested to carry either an OH- or an

O*-ligand and additional Cys-ligation [73], the measured data shows an unexpected upshift of about 2 cm^{-1} of the corresponding band at 871 cm^{-1} to 873 cm^{-1} (see Figure 60) [73]. Normally the exchange of a proton to the heavier deuterium should decrease the vibrational frequency of the corresponding stretching mode. Therefore, the shift could represent an artifact, as will be discussed in section 12.2.2., where this will be compared to computational data. If so, it would rule out a possible OH-ligation in the desulfo form.

12.1.3. Procedure for evaluation of ligation-sphere of molybdenum, as well as mechanistic properties of the Moco with QM-methods

For complete and detailed information on computational methods used in part V of this work, please consult section 5.3.

The following work uses the results from vibrational experiments to compare it to extensive calculations of different models of the Moco with different ligands (Cys; S*, SH, O*, OH) at the Mo-site (in different oxidation states: +VI, +V, +IV). In detail, the geometry of different Moco-models (see section 12.1.4. for more details on used models, as well as section 5.3.) has been optimized (using PCM/ε=4, if not stated otherwise) and the resulting structures used for subsequent Raman-spectra computations. These have been used to assign the bands in the experiments (for oxidized and reduced *Rc*FDH) and also to distinguish between different ligand-binding scenarios: The models and their vibrational spectra, which showed the best agreement to the measurements, were proposed to represent the correct ligation-sphere of the metal (in its respective oxidation state) and being a possible mechanistic intermediate in the catalytic oxidation of formate by the *Rc*FDH.

In chapter 12 (sections 12.2. and following) different models with Mo in its +VI-form and with different ligands (S*, SH, O* or OH; as well with or without ligated Cys) will be used for Raman-spectra computations and will be compared to the experimental subtraction RR-spectra of the oxidized *Rc*FDH in either active or inactive prestate form [73]. Furthermore, models with MoIV will be used in a similar way (S*-, SH-, O*- or OH-ligation; without ligated Cys) to determine the coordination-sphere of the metal in its +IV form, by comparison of computational results to experimental RR-spectra of the reduced *Rc*FDH [73].

Chapter 13 will then present results concerning the possible ligation of Cys to MoIV in the Moco, as well as possible intermediate formation of disulfide-bridge between sulfido-ligand and Cys in both MoVI- and MoIV-forms of the Moco. For this purpose several models and subsequent Raman-spectra will be computed and compared to the respective experimental vibrational analysis. Additionally, different models concerning MoV-structures are used and compared to reduced RR-measurements. To answer the question, whether or not formate binds directly to the metal-center, bond-breaking enthalpies for different models are discussed as well. Finally all findings are summarized in a new mechanism-proposal at the end of this chapter.

12.1.4. Preparing QM-methods: Testing of model size and method accuracy

Since the complete Moco is very large it is computational costly to use it for geometry optimizations, as well as Raman-spectra calculations. It will take long times for such molecules to converge, if at all. On the other hand using the complete Moco will render the most accurate results. Reducing the size of the Moco in the used models will make the computations less costly, but would also yield less accurate results. Big models will only allow few calculations, but with high accuracy, while smaller models enable more computations in the same time, but also with smaller precision. The question is, how much a size-reduced Moco-model differs in correctness, compared to the

complete Moco. Therefore, models with different sizes were tested in DFT-based QM geometry optimizations and subsequent Raman-spectra computations and the results compared to each other. Furthermore, reducing the convergence-criteria in the SCF QC calculations from 10^{-5} to 10^{-4}, as well as using a Newton-Raphson method for geometry optimization and its influence on the obtained vibrational frequencies will be discussed. This is due to the fact that some computations (as shown in section 5.3., as well as in 12.1.4.2.) only converge with this "hybrid" approach, named QC/Conver method. All tests have been carried out in vacuum, due to lesser computational costs.

12.1.4.1. Tests of influence of model size on preciseness of computed vibrational frequencies

Four different model types have been tested, varying in their respective size (s. Figure 61). All models include Mo , centered between two dithiolene moieties which are further connected to an pterin-unit each. MoVI was further ligated to either S, SH, O or OH and Cys (as denoted by X and Y, respectively in Figure 61). The same goes for MoIV, with exception of missing Cys and one vacant coordination site at molybdenum. This smallest model is named model 1. Model 1 excludes GDP-units (s. Figure 61, left, upper corner). Model 2 includes one CH_2 and one PO_4-group (in which the ending O is saturated by H; s. Figure 61, right, upper corner) for each dithiolene-ligand-side of the Moco (s. Figure 61, right, upper corner), while model 3 only excludes the guanine moieties (s. Figure 61, right, lower corner). The complete Moco is described by Model 4 (s. Figure 61, left, lower corner). Concerning naming, MoZ_X denotes Mo in oxidation-state +IV, +V or +VI and ligand S*, SH, O* or OH.

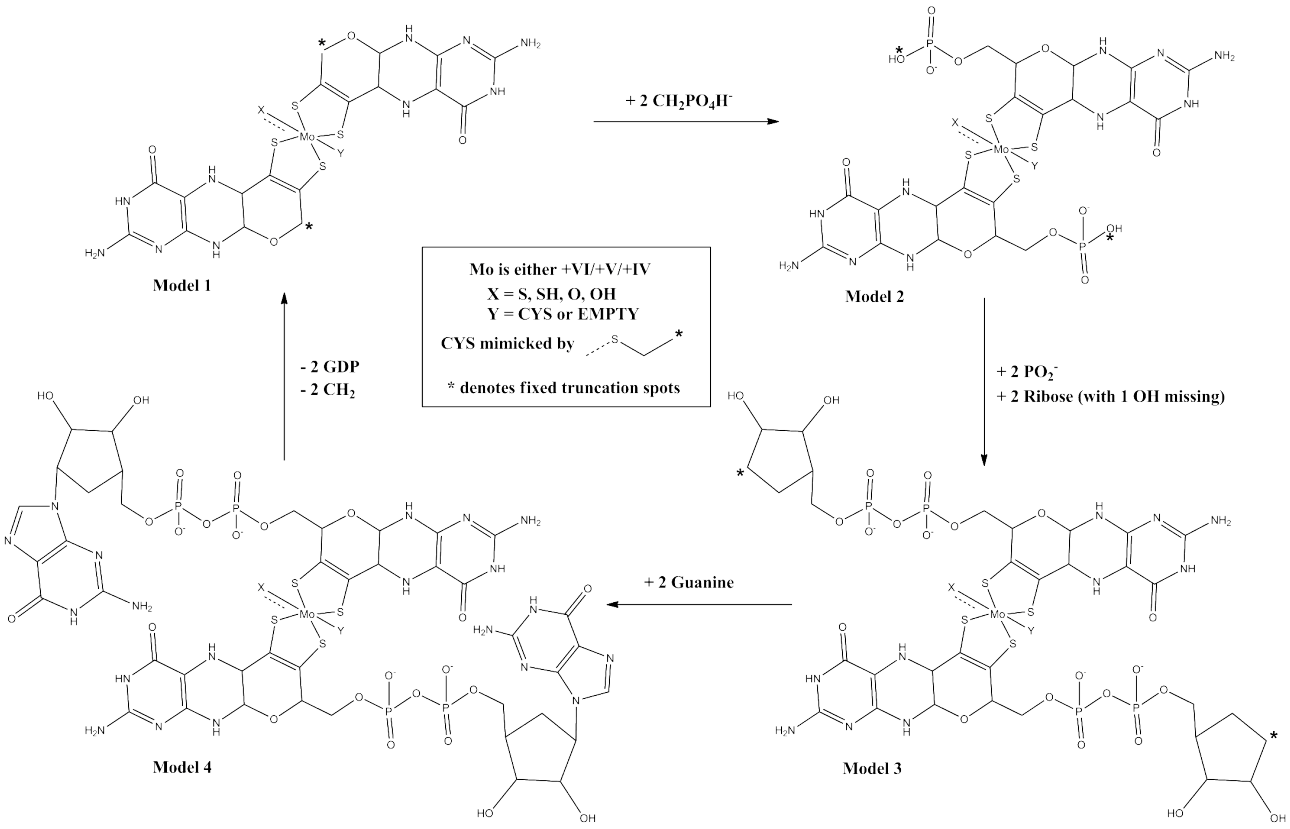

Figure 61. Schematic representation of used models 1-4 for QM; X denotes either S*, SH, O* or OH ligands; Y denotes either ligated Cys (mimicked by truncated version) or empty ligation site (as assumed for MoIV); Mo is in +VI, +V or +IV state; asterisk denotes fixed truncation spots

If used, in all four models, Cys has been mimicked by a truncated version, in which C_α has been

replaced by a CH_3 group which has been kept fixed (to simulate the lesser degrees of freedom of the residue in the protein matrix). Truncation spots (Model 1: C at cut between pterin and CH_2-linkage to PO_4; Model 2: O at cut between first and second PO_4-group; Model 3: cut between ribose and guanine) have been kept fixed during the geometry optimizations. In all models, the charges have been assigned as follows: each dithiolene (-2), deprotonated Cys (-1), doubly estered PO_4 (-1), O*/S*(-2), OH/SH (-1) and Mo either (+6), (+5) or (+4). Multiplicity has always been singlet for MoIV and MoVI models and doublet for MoV models. For more details on methods please see section 5.3. Following geometry optimization, frequencies and Raman-activities have been computed using the NMA approach [63],[156],[157],[158].

Extensive tests of models 3 and 4 (with different ligands O*, OH, S* or SH, as well as MoIV or MoVI) revealed difficulties in convergence. All tested model 4 compounds did not converge (independent of chosen DFT-method). Model 3 compounds converged only by using quadratically convergence methods (Newton Raphson algorithm; QC), while simultaneously lowering the convergence criteria from 10^{-5} to 10^{-4}. Such converged computations further showed problems in polarizability and Raman-activity estimations. Similar problems have already been faced upon addition of two phosphates to model 1 compounds. Therefore, the results are not shown here (s. Appendix A for more information about results for model 3 compounds and comparison to results for model 1). Contrasting this, model 1 compounds converged mostly with conventional SCF methods. Thus, it was important to test the accuracy of model 1 compound in comparison to larger ones (models 2-4), especially the influence of the PO_4 groups on the vibrational spectra. Therefore, Raman-spectra for models 1 and 2 have been computed and compared to each other: Here, such test will be demonstrated for model 1 and 2 compounds with MoIV and absent Cys ligand and ligated O-group at the metal center, since the other model 1 and 2 compounds yielded similar results regarding their accuracy and comparability to each other. Convergence has been achieved by standard SCF procedure. In Table 12 the important frequencies (marker bands as pointed out in section 12.1.2.) are shown, while in Figure 62 both spectra for models 1 and 2 are depicted.

The stretching vibrations corresponding to the potential ligands X (O*, OH, S* or SH) at the metal-site occur around 400 cm^{-1} (for S* and SH) and around 900 cm^{-1} (O* and OH). In the region in between (500 cm^{-1} to 800 cm^{-1}) mostly ring vibrations of the pterin ligands can be seen. And since the main interest lies in the determination of the ligation sphere of Mo, the comparison of the regions around 400 cm^{-1} and 900 cm^{-1} was determined to be sufficient:

The analyzed vibrations occur at similar frequencies in both models with small differences ($\Delta\nu$) between both Raman-spectra: ν(Mo-S): $\Delta\nu$=6 cm^{-1}; δ(S-Mo-S): $\Delta\nu$=13 cm^{-1}; δ(S-Mo-O*): $\Delta\nu$=1 cm^{-1}; ν(Mo-O*): $\Delta\nu$=2 cm^{-1}. Especially the Mo-O* stretchings are very near to each other in both models. Figure 62 shows also that in the so-called marker band regions (300-500 cm^{-1} and >800 cm^{-1}) the other bands are very similar in frequencies. Therefore, it was concluded that model 1 (excluding GDP's) was accurate enough for further computations and that the influence of phosphate, although detectable, is, for the purpose of analyzing the direct coordination sphere of molybdenum, negligible. In all following presented results, model 1 compounds have thus been used.

	ν (Mo-S) (cm^{-1})	**δ (S-Mo-S) (cm^{-1})**	**δ (S-Mo-O*) (cm^{-1})**	**ν (Mo-O*) (cm^{-1})**
MoIV_O model 1	342	364	315	938
MoIV_O model 2	336	351	316	936

Table 12. Vibrational frequencies of tested pentacoordinated models MoIV_O* with and without PO_4-units with standard SCF procedure; carried out in vacuum

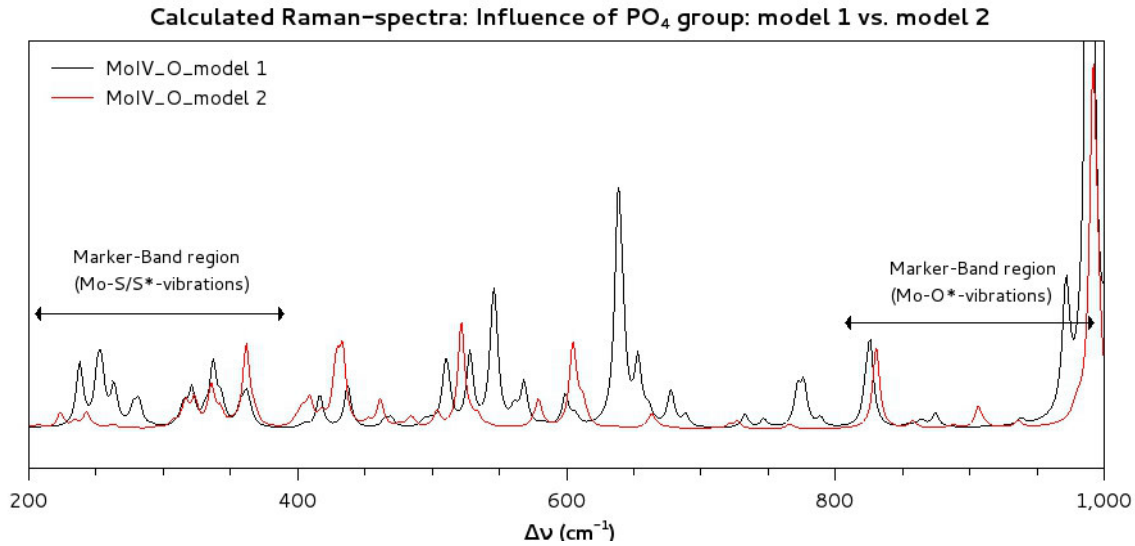

Figure 62. Calculated Raman-spectra for tested pentacoordinated models MoIV_O*: Model 1: red; Model 2: black; carried out in vacuum

12.1.4.2. Tests of influence of convergence method on preciseness of computed vibrational frequencies

Since some computations only converge using the QC-hybrid method (Newton-Raphson convergence algorithm, combined with lowering of convergence-criteria in SCF computations from 10^{-5} to 10^{-4}; see section 5.3; named QC/Conver method), the influence of it on the marker-band regions in model 1 compounds has been analyzed. Computation for this model 1 compound MoIV_O* and absent Cys has been carried out on one hand with standard SCF convergence procedure and the other with the hybrid QC-method and the resulting frequencies have been compared to each other. In Table 13 the corresponding results are depicted, while in Figure 63 the Raman-spectra for both models are shown.

The resulting marker band vibrations are very similar to each other in both calculations: ν(Mo-S): $\Delta\nu$=6 cm^{-1}; δ(S-Mo-S): $\Delta\nu$=17 cm^{-1}; δ(S-Mo-O*): $\Delta\nu$=10 cm^{-1}; ν(Mo-O*): $\Delta\nu$=3 cm^{-1}. Also the less-important middle region shows high similarity between both approaches, in terms of frequencies and intensity distributions. As shown in Figure 63, both spectra are almost identical. Therefore, it was concluded that the usage of the QC/Conver method for computations (should the necessity arise) is accurate enough. Similar results were found for tests with other models (not shown here).

	ν (Mo-S) (cm^{-1})	**δ (S-Mo-S) (cm^{-1})**	**δ (S-Mo-O*) (cm^{-1})**	**ν (Mo-O*) (cm^{-1})**
MoIV_O model 1 (standard SCF)	342	364	315	938
MoIV_O model 1 (QC)	336	347	325	935

Table 13. Vibrational frequencies of tested pentacoordinated MoIV_O*-models with standard SCF procedure and quadratic convergence method (with convergence accuracy at 10^{-4}) – QC/Conver method; carried out in vacuum

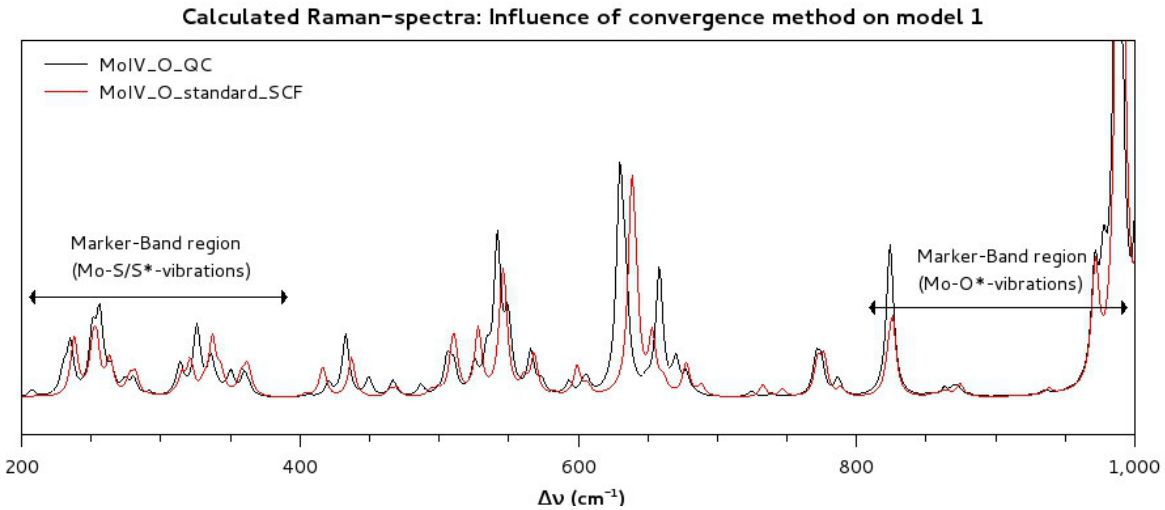

Figure 63. Calculated Raman-spectra for tested pentacoordinated MoIV_O*-model 1 compounds with standard SCF procedure (red) and quadratic convergence method (black; with convergence accuracy at 10^{-4}); carried out in vacuum

Conclusively, an effect of the size of the Moco on the corresponding computed Raman-spectra could be seen, while the choice of tested methods did not led to significant differences. The size-effect is only very small in the marker-band region which is why, it is of no concern for the investigation of possible direct ligands for MoVI, MoV and MoIV. If not stated otherwise, all following computations were carried out using model 1 compounds. Since in native environments, the protein would surround the Moco and influence vibrational motions, all calculations were done using a dielectric field constant ε=4 (so called implicit protein environment), as accepted as a good approximation for the electric field of a protein [198],[199]. Comparisons with computations in vacuum (ε=1) are shown in the Appendix B. There it can be seen that spectra-calculations with ε=4 yield better results than with ε=1, compared to experimental RR-data.

12.2. Refinement of cofactor structure of oxidized MoVI-form

This section focuses on the investigation of the coordination-sphere of MoVI in the Moco. Therefore, several model 1 compounds (as explained in crucial details in section 5.3.1. and in 12.1.4.) with different ligands X and Y (according to Figure 61) and maintained dithiolene-ligation are used for QM-geometry optimizations and subsequent vibrational frequency assignment. These results will be compared to experimental RR-data. This comparison will be used to validate the correct ligation-site of MoVI in both the catalytic active and inactive (prestate) forms.

12.2.1. Wild type (active; regarding S*- and Cys-ligation)

It has never been completely proven what ligand is coordinated directly to the MoVI-center in addition to the dithiolene moieties. In this section the analysis is limited to vibrations concerning a possible S*-ligation (frequency around 400-500 cm^{-1}) which will be compared to the subtraction RR-spectrum of the MoVI WT form (since here a band at 439 cm^{-1} is visible; s. Table 14 and section 12.1.2. for previous discussion of the probability of the presence of MoVI-S* vibration). Possibility of O*- and OH-ligation to MoVI will be discussed in section 12.2.2.

S*-ligation

To verify a sulfido-ligation at the MoVI-site, four different hexacoordinated model 1 compounds

with O*-, OH-, S*- and SH-ligand (and each model containing a ligation of Cys at MoVI) have been calculated and the vibrational frequencies have been subsequently computed with the NMA approach [63],[156],[157],[158] and compared to the experimental data [73]. Hexacoordination of MoVI has been chosen according to recent publications: The X-ray structure of FDH of *E. coli* [56] in its oxidized form showed a hexacoordination of the Moco. Simlarly the homology model studies by Hartmann et al. [49] used a hexacoordinated MoVI in the Moco. The results of the calculations can be seen in Table 14 and Figure 64.

	ν (Mo-S) (cm^{-1})	**ν (Mo-Cys) (cm^{-1})**	**δ (S-Mo-S) (cm^{-1})**	**δ (Mo-O-H) (cm^{-1})**	**δ (Mo-S-H) (cm^{-1})**	**ν (Mo-S*) (cm^{-1})**	**ν (Mo-O*) (cm^{-1})**
Exp. Subtr. RR (WT)	?	?	?	-	?	439	-
Exp. Subtr RR (prestate)	?	?	?	?	-	-	830/871
MoVI__O* (model 1 + ε=4)	337/365	347	232/332	-	-	-	874
MoVI__OH (model 1 + ε=4)	377/383	~250	130/189	767/786	-	-	622
MoVI__S* (model 1 + ε=4)	332/353	330/346	111/211	-	-	441	-
MoVI__SH (model 1 + ε=4)	354/376	278/346	183	-	640	354	-

Table 14. Vibrational frequencies for marker bands for hexacoordinated model 1 compounds with MoVI, differing in ligand (O*, OH, S* or SH); with usage of implicit protein environment (ε=4); compared to experimental subtraction RR-spectra (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

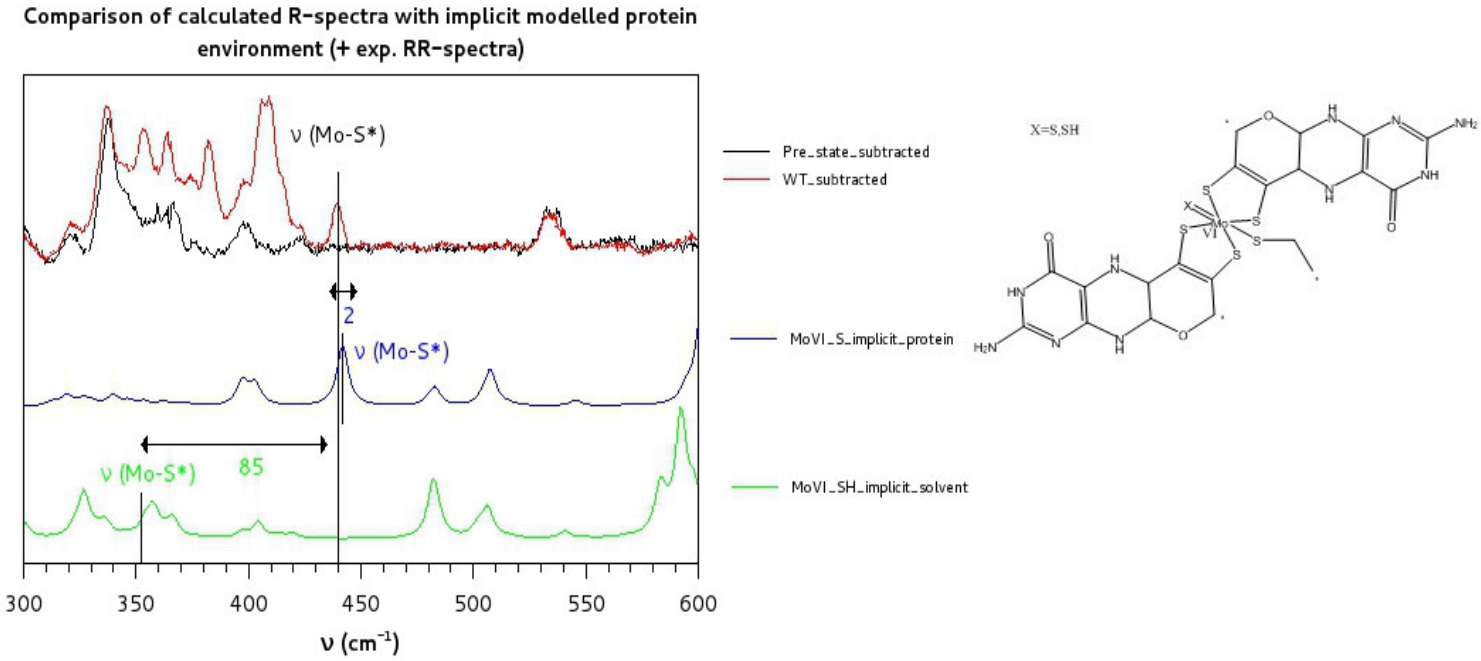

Figure 64. (left) Calculated Raman-spectra of hexacoordinated model 1 for MoVI_S* (blue) and MoVI_SH (green) (ε=4) and comparison to exp. subtraction RR-spectra: red: WT, black: desulfo-form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73]); (right) schematic visualization of used MoVI_S* and MoVI_SH model 1 compounds

The middle region (500-800 cm^{-1}; see Appendix B for Raman-spectra with full range from 300 cm^{-1} to 1000 cm^{-1}) of the computed spectra consists of mostly ring deformations and ring stretchings of the pterin ligands, while the Mo-S and Mo-Cys stretchings and the S-Mo-S deformations (for the dithiolene) can be found in the region 300-400 cm^{-1} (see Table 14). Mo-O* vibrations occur around 800 cm^{-1}, while the Mo-S* stretching motion can be seen between 350 cm^{-1} and 450 cm^{-1} (s. Table 14).

Since in the spectra of model 1 compounds with O*- or OH-ligand no band can be seen at frequencies around 450 cm^{-1} (s. Appendix B), which is visible in the RR spectrum of the WT, only the S*- and SH-containing models will be compared to the experimental subtraction RR-spectrum of the WT (depicted in Figure 64):

Very significant differences between both calculated spectra are visible: In the MoVI_SH model the Mo-S* stretching occurs at 354 cm^{-1}, while the Mo-S*-H bending motion is at 640 cm^{-1} and are thus strongly shifted from the experimental band observed at 439 cm^{-1} in the WT-spectrum (differences: $\Delta\nu$=85 cm^{-1}; $\Delta\nu$=201 cm^{-1}, respectively). In the MoVI_S* model we have a pronounced band at 441 cm^{-1} (corresponding to the Mo-S* stretching vibration) which is in very good agreement to the experimental band in the WT-spectrum ($\Delta\nu$=2 cm^{-1}). Also in the computed spectra of model 1 compounds with O*- and OH-ligands no band could be observed at the position around 439 cm^{-1}. Similar results have been obtained for computations in vacuum (ε=1; s. Appendix A).

But, it should be mentioned that a small signal can be seen around 400 cm^{-1} in all computed Raman-spectra of hexacoordinated models with SH-ligation present at the Moco. In Figure 64 it appears at ~405 cm^{-1} and is about 34 cm^{-1} shifted from the observed band in the RR-spectrum at 439 cm^{-1}. This band (independent of the used model or environment; s. Appendix A and B for further Raman-spectra of different models) has always a lower activity than the Mo-S* stretching in the MoVI_S* models. Also it is always shifted to lower wavenumbers compared to the Mo-S* stretching in the Mo_S model. Furthermore, H/D-exchange experiments by Wahlefeld et al. [73] revealed no shift of the observed signal at 439 cm^{-1} and did not coincide with the calculated shift in computed Raman-spectra with MoVI-SD ligation (by Mroginski et al. [200]). This makes it very unlikely for a SH-liagtion to be present in the oxidized WT form of the Moco. Additionally, experimental studies all suggested a sulfido-ligation in the oxidized form of the Moco, in contrast to a thiol-group, to be present: E.g. XAS-analysis of *Rc*FDH [50], as well as X-Ray studies of FDH of *E. coli* [52] suggested a S*-ligation. Further studies, such as from Mota et al. [53] and Maia et al. [54], as well as from the group of Silke Leimkühler [48],[49] assumed also a sulfido-ligation.

Thus, computed Raman-spectra have been analyzed in terms of the position of a Mo-S* stretching relative to the experimental band. And it can be concluded that in the oxidized WT form of the Moco, a sulfido-ligand is directly coordinated to the MoVI-center (with present Cys-ligand), since the corresponding calculated spectra are in best agreement to experimental data concerning the vibration at 439 cm^{-1}.

Cys-Ligation

In the literature, it is assumed that a Cys ligand is present in the first coordination sphere of MoVI [50],[52],[53],[56],[182],[183]. The influence of Cys on the vibrational signature was investigated by four different pentacoordinated MoVI model 1 compounds with either O*-, OH-, S*- or SH-ligands, while no cysteine was present. The results have been compared to model 1 compounds with hexacoordinated MoVI and directly ligated Cys and to experimental data. In the following only the models with O*-ligation will be discussed, because they show the influence of the presence or

absence of Cys on the spectra in the best way (for more details about the other models, please see the Appendix B). The results of the spectra-computations of model MoVI_O* are shown in Table 15 and Figure 65.

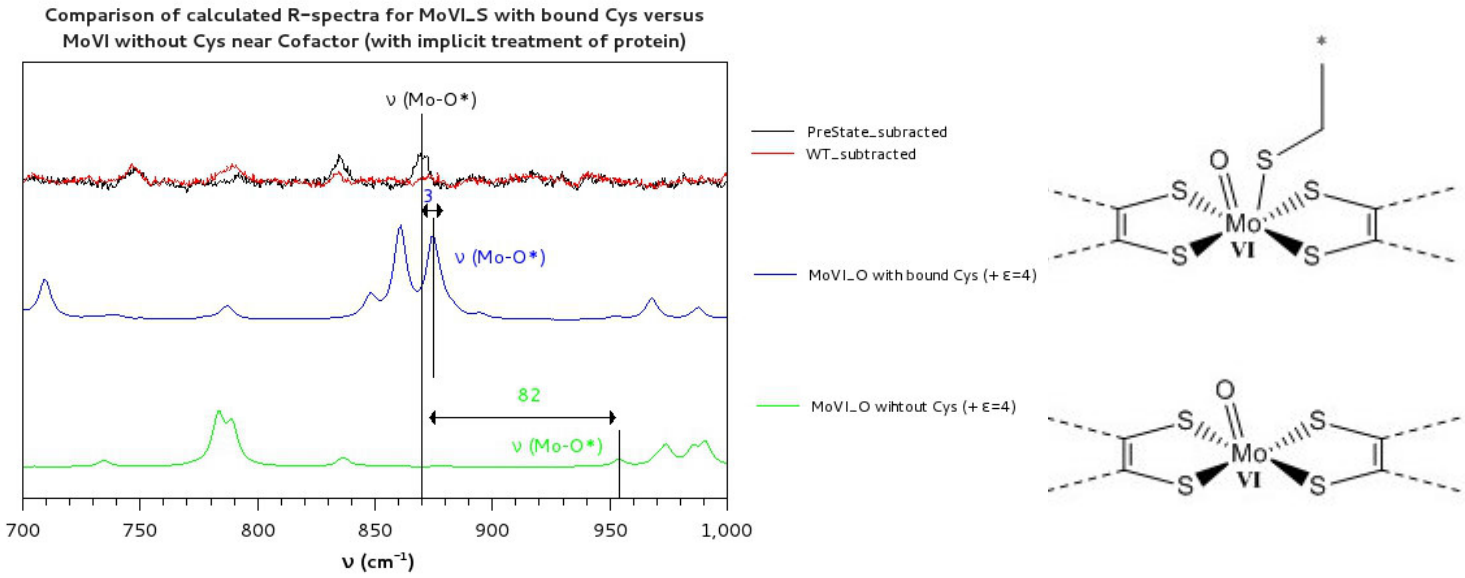

Figure 65. (left) Calculated Raman-spectra for MoVI_O* model 1 compounds (ε=4) with (blue) or without (green) Cys and comparison to experimental subtraction RR-spectra: red: WT, black: desulfo-form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73]); (right) schematic representation of MoVI_O* model 1 compounds with or without ligated Cys

	ν (Mo-S) (cm^{-1})	**ν (Mo-Cys) (cm^{-1})**	**δ (S-Mo-S) (cm^{-1})**	**δ (Mo-O-H) (cm^{-1})**	**δ (Mo-S-H) (cm^{-1})**	**ν (Mo-S*) (cm^{-1})**	**ν (Mo-O*) (cm^{-1})**
Exp. Subtr. RR (WT)	?	?	?	-	?	439	-
Exp. Subtr RR (prestate)	?	?	?	?	-	-	830/871
MoVI__O* with bound Cys (model 1 + ε=4)	337/365	347	232/332	-	-	-	874
MoVI_O* without Cys (model 1 + ε=4)	361-382	-	176	-	-	-	953

Table 15. Vibrational frequencies for marker bands of model 1 compounds MoVI_O* without Cys and with usage of implicit protein environment (ε=4); compared to experimental subtraction RR-spectra (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

The model with metal-bound Cys shows a band at 874 cm^{-1} (Mo-O* stretching; see Table 14), which is in good agreement to the band at 871 cm^{-1}, present in the experimental RR-spectrum of the inactive desulfo-form of the oxidized species which has been assumed to be due to a Mo-O* stretching vibration typical for this region (and because a MoVI-O* ligation has been found in XAS-experiments [50]). The difference between exp. and computed frequency is $\Delta\nu$=3 cm^{-1}. In contrast to this, the spectrum of the MoVI_O* model without Cys shows no band around 870 cm^{-1}. The ν (Mo-O*) appears at 953 cm^{-1} and is thus strongly shifted compared to the positions of this

band in the RR- and the computed Raman-spectrum ($\Delta\nu$=82 cm^{-1}). A similar observation can be made for the MoVI_S* models (not shown here): The hexacoordinated model with metal-bound Cys yields the Mo-S* stretching only $\Delta\nu$=2 cm^{-1} shifted from the experimental band at 439 cm^{-1}, while the pentacoordinated model without Cys shows a displacement of this vibration of about $\Delta\nu$=80 cm^{-1} compared to the experiment. This was also found in computed Raman-spectra for model 1 compounds in vacuum (s. Appendix B).

In conclusion, a missing Cys-residue ligated to MoVI leads to strong displacements of marker bands (e.g. ν (Mo-S*) and ν (Mo-O*)) in the corresponding computed Raman-spectra and thus bad agreement to experimental RR-data. This disagreement is in contrast to the findings presented in sections 12.2.1. and 12.2.2. and to experimental publications which suggest a sulfido-coordination in the active WT species with oxidized Moco and a oxo-coordination at MoVI in its inactive prestate form, while present Cys-ligation at MoVI [50] (s. sections 12.1.1.1. and 12.2.1. and 12.2.2.). Inclusion of directly coordinated Cys-ligand in the calculated models leads to a good representation of RR-spectra and is consistent with experimental and theoretical findings. Therefore, Cys is most likely directly ligated to MoVI and is part of the coordination sphere. The Moco in the oxidized and active WT form of *Rc*FDH is thus carrying a hexacoordinated MoVI, ligated with two dithiolene-moities, one S*-ligand and one native Cys-ligand.

12.2.2. Prestate (inactive desulfo form)

In the absence of the FdsC, an inactive form of the oxidized *Rc*FDH (prestate or desulfo-form) exists which is presumably desulfurized: As in *E. coli*, the FdsC equivalent gene may be responsible for sulfuration of the cofactor [50]. Therefore, it remains questionable what group is ligated to the MoVI atom, if not a sulfido-group. XAS-studies by Schrapers et al. [50] found a Mo-O* bond to be present in the prestate and the absence of the Mo-S* ligation. This led to the suggestion that a oxo-group is ligated to MoVI (instead of a sulfido) in the inactive form of *Rc*FDH.

O*- vs. OH-ligand

To determine the possible ligation partner, the calculated Raman-spectra for the model 1 complexes (as pointed out in section 12.2.1.) with ligated Cys have been compared to the corresponding experimental subtraction RR-spectrum of the desulfo-form. In this an exclusive band at 871 cm^{-1} is visible which does not occur in the WT-spectrum and has been assumed to be attributed to Mo-O* vibrations, typical in this region. Further, this band should correspond to a vibration exclusive for the prestate. Thus, this will be the marker-band for comparison to the computed Raman-spectra. The calculated spectra and frequencies are shown in Table 14 and Figure 66.

Although the hexacoordinated MoVI_S* model 1 compound does show bands near 870 cm^{-1}, it also shows signals around 440 cm^{-1} which cannot be observed in the experimental RR-spectrum of the prestate. That is why S*-ligation can be ruled out with MoVI in the prestate-Moco. SH-ligation can be ruled out by H/D exchange [73] (as discussed later) and the appearance of bands near 400 cm^{-1} and 480 cm^{-1} which are not visible in the RR spectrum. Raman-spectra for hexacoordinated model 1 compounds with O* and OH ligated to MoVI on the other hand, show similarities to the experiment in the region ~800-900 cm^{-1} and no prominent bands ~440 cm^{-1}. Therefore, the computed Raman-spectra of the two latter model 1 compounds will be discussed in further detail and compared to the RR-experiments.

As can be seen in Figure 66 and Table 14 the MoVI_O* model yields a signal for the Mo-O* stretching in the corresponding calculated Raman-spectra at 874 cm^{-1} in very good agreement to the

experimental band at 871 cm^{-1} (shifted about only $\Delta\nu$=3 cm^{-1}). A second band appears in near proximity at ~860 cm^{-1}, in agreement to the RR-band at 830 cm^{-1} ($\Delta\nu$=30 cm^{-1}), thus resembling the two-band motif of the RR-spectrum. In contrast to this, in the MoVI_OH model, the nearest oxo-containing band is the Mo-O*-H bending motion at 776 cm^{-1} (shifted away from the exp. of about $\Delta\nu$=95 cm^{-1}) only poorly visible and also not featuring the double-band motif seen in the RR-spectrum. The Mo-O* stretching motion is at 622 cm^{-1} which displays a strong displacement from the experiment (of $\Delta\nu$=249 cm^{-1}) and is therefore negligible. The small signals visible around ~850 cm^{-1} in the Raman-spectrum of the MoVI_OH model are due to ring vibrations of the pterin moieties and will not be considered as equivalent for the experimental signal at 871 cm^{-1}, because this band should represent a vibration including oxygen. Also further H/D exchange experiments ruled out the possibility of an OH-ligation (see next paragraph for more details). Thus, the model with O*-ligation to the metal is in best agreement to the experiment.

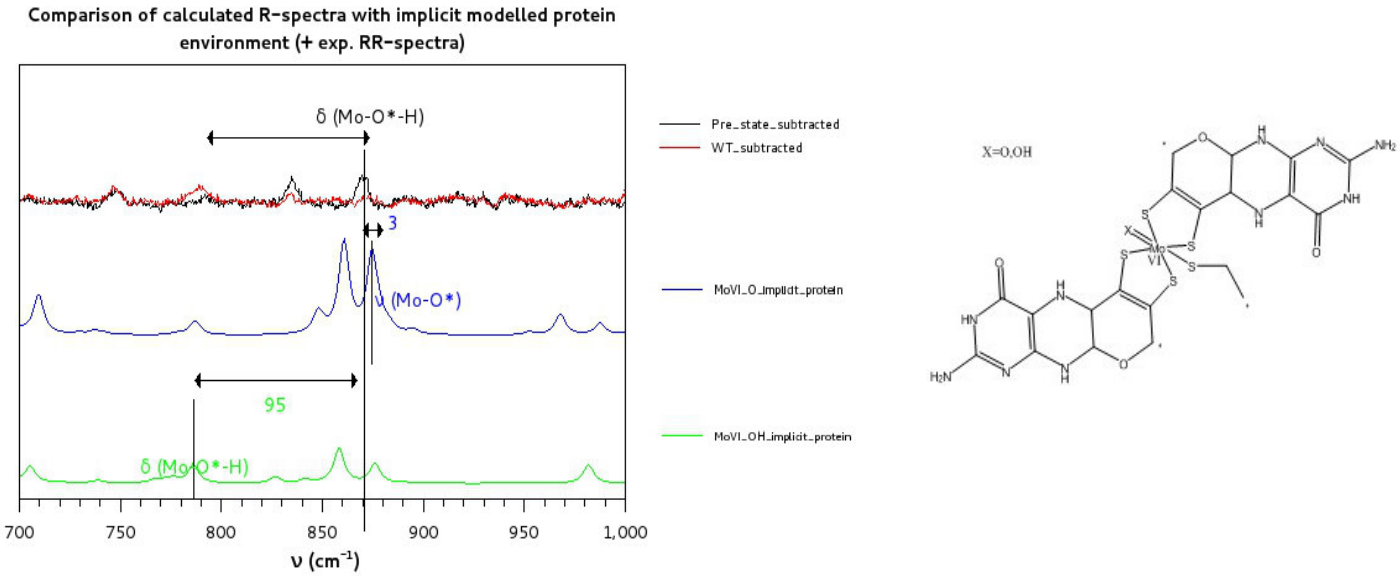

Figure 66. (left) Calculated (ε=4) Raman-spectra for MoVI_O* (blue) and MoVI_OH (green) hexacoordinated model 1 compounds and comparison to experimental subtraction RR-spectra: red: WT, black: desulfo-form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73]); (right) schematic visualization of used MoVI_O and MoVI_OH model 1 compounds

Also, in Raman-spectra calculated in vacuum or with model 3 compounds (see Appendix A and B), the Mo-O* vibration in the MoVI_O* model is always in best agreement to the experimental one at 871 cm^{-1}. Therefore, it can be concluded that the missing FdsC gene indeed leads to the incorporation of an oxo-ligand in the first coordination sphere instead of the sulfido-one in the active form, leading to an inactive hexacoordinated Moco (with two dithiolenes, one Cys and one oxo-group ligated to the MoVI).

H/D exchange

To further verify these findings, H/D-exchange experiments have been taken out by Stefan Wahlefeld [73] (see section 12.1.2.) and compared to computations. The idea has been that if an OH-ligand would be present at the MoVI-site in the prestate of *Rc*FDH, the deuteration of the cofactor would also lead to an exchange of the proton of the OH-group. This would further lead to a measurable shift of the Mo-O* stretching, i.e. the band at 871 cm^{-1} in the corresponding RR-

spectrum. Experimental shifts, then should be reproducible by the computed Raman-spectrum of a MoVI_OD model 1 compound. Here the exchangeable proton of the OH-group in a hexacoordinated MoVI_OH model 1 compound and metal-bound Cys has been replaced by a deuterium and the geometrically optimized system has been used for subsequent Raman-spectra computations. The obtained results have been compared to the experimental findings (see Figure 60) and are shown in Table 16.

	δ (Mo-O-H) (cm^{-1})	**Δν (cm^{-1})**
Exp. Subtr RR (prestate)	830/871	
Exp. Subtr RR (prestate) → H/D exchange	830/873	+2
MoIV_OH model 1 (bound Cys) + ε=4	767/786	
MoIV_OD model 1 (bound Cys) + ε=4	566/627	-201/-159

Table 16. Calculated vibrational frequencies for hexacoordinated MoVI_OH and MoVI_OD model 1 compounds (ε=4); shifts induced by deuteration denoted by Δν; comparison with experimental RR-spectra (H/D exchange; 1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

In the RR-experiment only an upshift of 2 cm^{-1} of the band at 871 cm^{-1} could be observed [73]. In contrast to this, in the calculated spectrum of the MoVI_OD model, this cannot be reproduced (see Table 16). The Mo-O* stretching is at 627 cm^{-1}, showing a small upshift of about Δν=5 cm^{-1} compared to the computed Raman-spectrum of the MoVI_OH model, but is too far away from to the experimental band at 871/873 cm^{-1} and thus negligible. The Mo-O*-H vibrations are downshifted to 566/627 cm^{-1} of about Δν=201/159 cm^{-1} compared to the undeuterated model. They are now in even poorer agreement to the experimental bands at 830/873 cm^{-1} than in the MoVI_OH model. No other band is visible around 870 cm^{-1} in the MoVI_OD model. The same was found for computations in vacuum (s. Appendix B). Since the experimentally observed upshift could not be reproduced with the MoVI_OD model, in *argumentum e contrario*, an OH-ligand is not present in the desulfo-form.

In conclusion, an O*-ligation at the hexacoordinated MoVI-site in the inactive form of *Rc*FDH is present and the measured band at 871 cm^{-1} is due to the Mo-O* stretching motion. This is further underlining the necessity of the FdsC gene for the incorporation of the sulfido-group and the correct catalytic activity of the *Rc*FDH (see also [48],[49]). Additionally, this shows the importance of the sulfido-group in the catalytic process which will be discussed in sections 13.6. and 13.7.

12.3. Refinement of cofactor structure of reduced MoIV-form

The Moco of *Rc*FDH exists in at least two (possibly three) different oxidation-states, concerning the molybdenum. Mo can be either oxidized (MoVI) or reduced (MoIV and possibly MoV). No inactive desulfo form for either reduced form has been reported yet. Similar to the oxidized form of *Rc*FDH, it is widely assumed that the reduced MoIV carries a S*-ligand in addition to two dithiolene moieties [48],[49],[50],[52],[53],[54],[55]. In contrast to this, the reduced form of the metal has been assumed to be only pentacoordinated, without Cys to be part of the direct coordination sphere. This was shown in X-ray analysis of *E. coli* [52] by studies of Hartmann et al. [49] and theoretical investigations of Mota et al. [53] who showed the strong electrostatic repulsion between MoIV and Cys in the reduced state of the FDH. Therefore, in the following computated models no Cys has been present. The possible influence of Cys in the reduced form of the Moco is further discussed and analyzed in section 13.3.1.

Several RR-spectra of the reduced form of *Rc*FDH have been measured, all looking very much alike and also very similar to the ones for the oxidized forms of the enzyme (s. section 12.1.2.). It is difficult to say with absolute certainty, if the obtained bands are purely from the reduced enzyme. Possibly the measured species was a mixture of both oxidized and reduced protein. All measured spectra show a signal at ~439 cm^{-1} which may be due to a Mo-S* stretching (as already determined for the oxidized species) and thus lead to the hypothesis of the presence of an S*-ligation at the metal ion also in its reduced +IV form. Therefore, in the following the analysis of the computed results will focus on the comparison of calculated signals with this experimental band at ~439 cm^{-1} .

S*-ligation

To verify S*-ligation (in absence of Cys) in the +IV form of the Moco/Mo, four different models have been considered, differing in the possible ligand at the MoIV-site: O*, OH, S*, SH. Furthermore, all models did not include a cysteine. As starting structures, the homology model of the reduced *Rc*FDH has been used [49]. The optimized geometries of these models have been used for subsequent Raman-spectra computations according to the NMA approach [63],[156],[157],[158] which have been compared to the experimental RR-spectra (see Figures 58 (right) and 59)

The spectra-composition is very similar to the one of the MoVI-model 1 compounds. The region between 500 and 800 cm^{-1} is mostly containing ring-deformations of the pterin-ligands, while between 800 and 900 cm^{-1} Mo-O* stretchings can be found (in the MoIV_O* model; see Appendix A and B for further informations on MoIV_O* and MoIV_OH model 1 Raman-spectra). All four Raman-spectra show similar band patterns in the region between 300 and 400 cm^{-1}, all contributing to either Mo-S (dithiolene) or Mo-S* (sulfido-group) vibrations (the latter only in the MoIV_S* and MoIV_SH models). The MoIV_O* model yields a significant big peak near 900 cm^{-1} which cannot seen in the experimental spectrum. OH-ligation can be ruled out by H/D-experiments [73], as discussed below. Furthermore, as shown for the oxidized (MoVI-) models, bands (concerning metal-ligand vibrations) appearing in near proximity to the measured one at 439 cm^{-1} can be found in the computed Raman-spectra of the MoIV_S* and MoIV_SH models. Therefore, these two are analyzed in more detail and the results are shown in Figure 67 and Table 17.

	ν (Mo-S) (cm^{-1})	**δ (S-Mo-S) (cm^{-1})**	**δ (Mo-S-H) (cm^{-1})**	**ν (Mo-S*) (cm^{-1})**
Exp. RR	?	?	?	439
MoVI_S* (model 1 + ε=4)	322-360	182	-	461
MoVI_SH (model 1 + ε=4)	344-386	175	519	364

Table 17. Vibrational frequencies for marker bands of different pentacoordinated MoIV model 1 compounds, differing in ligand (S or SH); usage of implicit protein environment (ε=4) and compared to experimental RR-spectrum (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73])

In the MoIV_SH model the Mo-S* stretching and the Mo-S*-H deformation at 364 cm^{-1} and 519 cm^{-1}, respectively, are in poor agreement to the band at 439 cm^{-1} in the RR-data (shifted away from the exp. bands of about $\Delta\nu$=75 cm^{-1} and $\Delta\nu$=80 cm^{-1}, respectively). Contrasting this, the Mo-S* vibration in the MoIV_S* model at 461 cm^{-1} is nearer to the experimental signal (about $\Delta\nu$=22 cm^{-1}). Its deviation of about 58 cm^{-1} from the Mo-S*-H deformation signal in the MoIV_SH model,

makes the MoIV_S* model a better candidate for the correct ligand-binding situation at MoIV in the Moco.

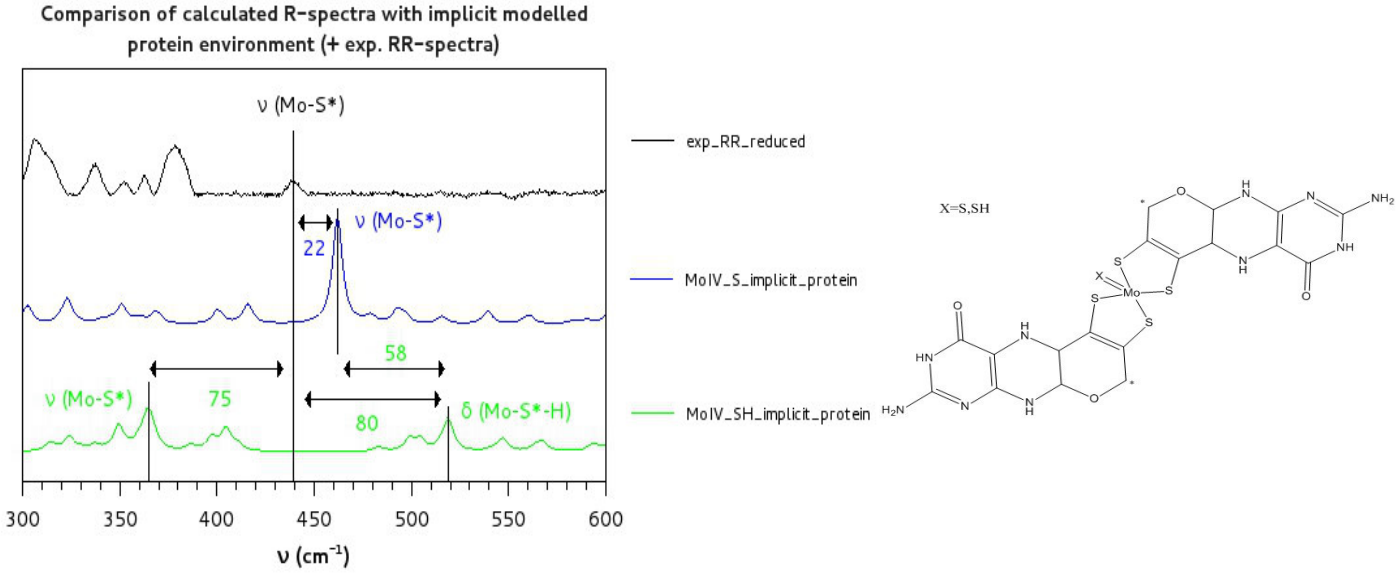

Figure 67. (left) Calculated Raman-spectra for MoIV_S* (blue) and MoIV_SH (green) pentacoordinated model 1 compounds (ε=4) and comparison to experimental RR-spectrum: (black: reduced form: 1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73]); (right) schematic representation of used MoIV_S* and MoIV_SH model 1 compounds

But it should be mentioned that the Mo-S* vibration in the MoIV_SH model at 364 cm^{-1} is very near to the experimental band at 378 cm^{-1}. Also, in all the models with MoIV-SH-ligation (also for computations of model 1 and 3 compounds in vacuum; see Appendix A and B), a signal near ~400 cm^{-1} can be observed which can be attributed to Mo-S stretchings, concerning the dithiolene-metal ligation. The observed signal at 439 cm^{-1} in the experimental reduced RR-spectrum is most likely due to a metal-sulfido vibration, but a signal near 378 cm^{-1} or 405 cm^{-1} can also be observed in several reduced RR-spectra (see Figure 58, right). These two bands and its agreement to computational data could indicate a contribution of a MoIV-SH state to the measured reduced species. This would indicate a mixed state, with either S*- or SH-ligation to MoIV, most probably depending on the time-step of the reaction cycle. H/D-exchange experiments [73] yielded no shift in frequencies of signals at 378 cm^{-1} and 439 cm^{-1}, but a very small shift of the band near 405 cm^{-1}. This could indicate that we have indeed a short-living (intermediate) MoIV-SH species in addition to the more stable MoIV-S* state. But since, the experimental band at 378 cm^{-1} could also come from a reduced FeS-species and the H/D-effects are only small, it is not yet clear. See sections 13.3.1. and 13.6. and 13.7. for further details and discussion of the possibility of the existence of a MoIV-SH ligation-scenario.

Conclusively, since the MoIV_S* model gives the best agreement to the RR-data (especially concerning the band at 439 cm^{-1} which can be assigned to the Mo-S* stretching motion), it is proposed that MoIV is coordinated to a sulfido-ligand, additional to two dithiolene units and without ligated Cys, yielding a pentacoordinated metal. This further underlines the importance of the S*-ligand for the catalytic cycle, since it is present in both active oxidized +VI and reduced +IV forms of the metal. Further studies of a possible Cys-ligation to MoIV are shown in section 13.3.1.

12.4. Discussion of refinement of cofactor structures

The investigations of the coordination spheres of both MoVI and MoIV revealed the most probable ligation situations at the metal-site in both +VI an +IV oxidation-states: A hexacoordination for the +VI form of the metal has been determined. A sulfido-ligand is bond to the metal in the active WT form of the enzyme in *Rc*FDH. Simultaneously, Cys is ligated to the oxidized metal, as well as two dithiolene moieties. Furthermore, a pentacoordinated +IV form of the metal has been assumed (due to experimental and theoretical findings) with two dithiolenes and a sulfido-group ligated to the MoIV as the most probable and stablest form. Additionally, the possibility of an intermediate SH-ligation to MoIV has been discussed.

This is in agreement with the literature: Crystal-structure analysis of FDH of *E. coli* suggested similar structures and ligations for both MoVI and MoIV [52]. Both Hartmann et al. [49] and Maia et al. [54] included a S*-ligation for MoVI and MoIV in their respective mechanisms. The latter also discussed the possibility of the formation of a SH-ligand in the MoIV state. An intermediate SH-ligation to a reduced molybdenum (MoIV or MoV) has also been discussed in other studies [50],[52],[54],[55],[196].

In contrast to the here presented results, Mota et al. [53] suggested a pentacoordinated active MoVI-form of the Moco, with S*-ligation and no direct binding of cysteine to molybdenum. But this structural proposal has been revised recently [54]. Furthermore, Boyingtons X-Ray analysis of FDH-H of *E. coli* [56] suggested an OH-ligation to MoVI which leaves the coordination-site upon reduction of Mo. This has been reinterpreted by Raajmakers et al. [52] who suggested a S- or SH-ligand present in both MoVI and MoIV, instead of Mo-OH-ligation. Furthermore, the observed Mo-X bond of about ~2.2. Å [56] is normally too long for a Mo-OH bond and is in the typical region of a Mo=S ligation and normally too short for a Mo-SH bond as well.

For the inactive prestate of the MoVI-form of the Moco an oxo-group has been determined to be ligated to molybdenum instead of a sulfido-group. In addtion a Cys-ligand is present at MoVI, as well as two dithiolene moieties. This form is also called desulfo-form. Studies by Hartmann et al. [49], as well as XAS-data by Schrapers et al. [50] also discussed the possibility of the existence of this hexacoordinated inactive species and in the latter an oxo-ligation to MoVI has also been assumed for the prestate.

Conclusively, the here depicted results determine the ligation-sphere of both MoVI and MoIV states *via* computational methods and for the first time in combination with vibrational spectroscopic analysis.

Next, the involvement of these structures in the catalytic oxidation of formate has to be investigated. Especially, the position of the Cys-residue throughout the reaction cycle and its possible displacement from the metal, as well as possible direct ligation of formate to molybdenum are topics for discussion in recent publications. Also the involvement of the suflido-ligand is not yet clear. These and further questions are subject of the investigations shown in the next section.

Chapter 13

Evaluation of mechanistic properties

The knowledge of the complete ligation sphere of the Moco in the MoVI and MoIV states, as determined *via* computational Raman-spectra and subsequent comparison with experimental RR-data (as provided by Stefan Wahlefeld [73]), provides the necessary basis for further investigations of the catalytic process at the Moco-site in *Rc*FDH.

13.1. State of art of the catalytic process

Different mechanisms have been proposed for the catalytic activity at the Moco of FDHs (s. section 3.3.), in which different properties are attributed to the Moco and the Cys. In the studies of Schrapers et al. and Hartmann et al. [49],[50] the dissociation of the Cys-ligand from MoVI in favor for direct coordination of formate has been suggested. Cys is then responsible for abstraction of a proton from formate [49]. This was contradicted by the investigations of Maia et al. [54] which suggested an unbound formate. They further concluded that formate was interacting with the S*-ligand *via* a hydride-transfer and the generation of a SH-ligand at MoIV [54]. Robinson et al. [55] suggested a mechanism with characteristics from both former ones: They proposed the dissociation of Cys from the metal and subsequent direct coordination of formate to molybdenum. Additionally, they included the possibility of a hydride-transfer from formate to the sulfido-group at the Moco. Mota et al. [53] discussed that the possible on-and-off-going of Cys might be mediated by the formation of a temporal disulfide linkage between S*- and Cys-ligand, whereas the sulfido-group stays ligated at the metal. Furthermore, the presence of a MoV-intermediate has been detected [50], [54],[196] suggesting a step-by-step electron-transfer from MoVI to MoIV and *vice versa*. A proton in the second coordination sphere of molybdenum and its interaction with MoV [54],[196] led to the suggestion of a possible SH-ligation, as also discussed by Schrapers et al. [50].

All these findings represent important possibilities for key-steps in the reaction cycle at the Moco. Since they in parts contradict each other, further investigations are necessary to create a clear image of the reaction mechanism. Especially the possible unbinding of Cys during the catalytic cycle is of great interest, since it is controversially discussed in the science community. The same is true for the role of both Cys and S*-ligand in the oxidation of formate and abstraction of hydrogen from it.

Therefore, in this chapter the ligation of Cys to MoIV states is investigated. Furthermore, the possibility of its dissociation from molybdenum *via* the formation of a disulfide-bridge is studied. Finally, the presence of a MoV-species and thus a stepwise reduction of Mo is analyzed. Together with the results presented in chapter 12, these investigations will help to create a new and improved mechanism for the catalytic oxidation of formate at the Moco-site.

Different model 1 compounds, regarding the ligation of Cys to MoIV and the formation of a S*-S(Cys) bond in both MoVI and MoIV, as well as concerning the coordination-sphere of MoV have been built. These have been geometrically optimized via DFT and used for subsequent Raman-spectra calculations. These have been compared to experimental RR-spectra [73] and used for determination of the correct ligand-assignment in the different redox-states. Furthermore, bond-dissociation enthalpies have been computed, to analyze the possible binding of Cys and $HCOO^-$ to MoVI and MoIV. Details about model-building can be found in the respective sections. Possible binding sites and modes of action for catalytic inhibitors (azide) to MoVI and its consequences for the reaction cycle, are also studied and discussed in Appendix C.

13.2. Disulfide-bridge formation and possible unbinding of cysteine in the MoVI state

To clarify the question, whether Cys is leaving the ligation site of Mo and whether this happens in the MoVI-state or the MoIV-state and also if this happens *via* temporary formation of a S*-S bond, different computations have been carried out: Two different model 1 compounds have been built (MoVI_S*-S and MoIV_S*-S; s. Figure 68), in which Cys is covalently linked to the sulfido-ligand at the metal center (a charge of -1 was applied for the S*-S(Cys)-ligand). The MoVI and MoIV models varied only in the redox state of the metal and the usage of a different starting structure based on the homology models of either the oxidized enzyme or the reduced one [49]. The models have been geometrically optimized with QM methods (as described in section 5.3.) and subsequent Raman-spectra have been computed using the NMA approach [63],[156],[157],[158]. The results for MoIV are discussed in 13.3.2.

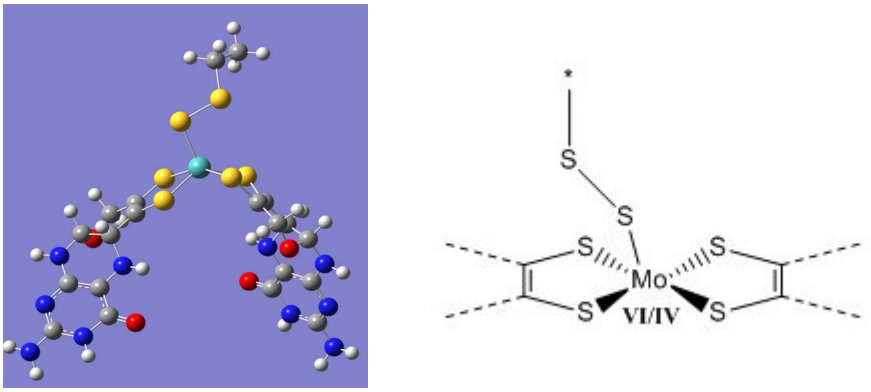

Figure 68. Representation of pentacoordinated MoVI/MoIV model 1 compounds with explicit treatment of disulfide bond between sulfido ligand and Cys-ligand: (left) MoVI_S_S_Cys model; (right) schematic representation

It had been assumed that the probes for vibrational measurements and thus the experimental RR-spectra might contain intermediate structures like the ones with a disulfide bridge at the Moco. The corresponding stretching vibration of the S*-S bridge, which also occurs in the region of 440 cm^{-1}, might be overlapping with the Mo-S* stretching at 439 cm^{-1} (s. 12.2.1. and 12.3.). Therefore, the computed Raman-spectra have been compared to RR-data with respect to the band at 439 cm^{-1}.

	ν (Mo-S*) (cm^{-1})	**ν (S*-S) (cm^{-1})**	**δ (Mo-S*-S) (cm^{-1})**	**δ (S*-S-C) (cm^{-1})**
Exp. Subtr. RR (WT)	439	?	?	?
Exp. Red. RR	439	?	-	-
MoVI_S*-S (model 1 + ε=4)	398	489	489/595	489/595/712
MoIV_S*-S (model 1 + ε=4)	388	423/585	423/585	423/757/946

Table 18. Calculated vibrational frequencies for pentacoordinated MoVI and MoIV model 1 compounds (ε=4) with explicit modeling of disulfide bond between sulfido and Cys-ligand (MoVI/IV_S*-S) and comparison to corresponding experimental RR-spectra (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 10 mM formate; provided by Stefan Wahlefeld [73])

Table 18 shows the results of the frequency calculations for both models. The results for the MoIV model with S*-S bond are discussed in section 13.3.2. The computed Raman-spectrum has been compared to the experimental RR-date of the WT of *Rc*FDH, as well as to the MoVI_S* model 1 compound (see Table 14) as depicted in Figure 69.

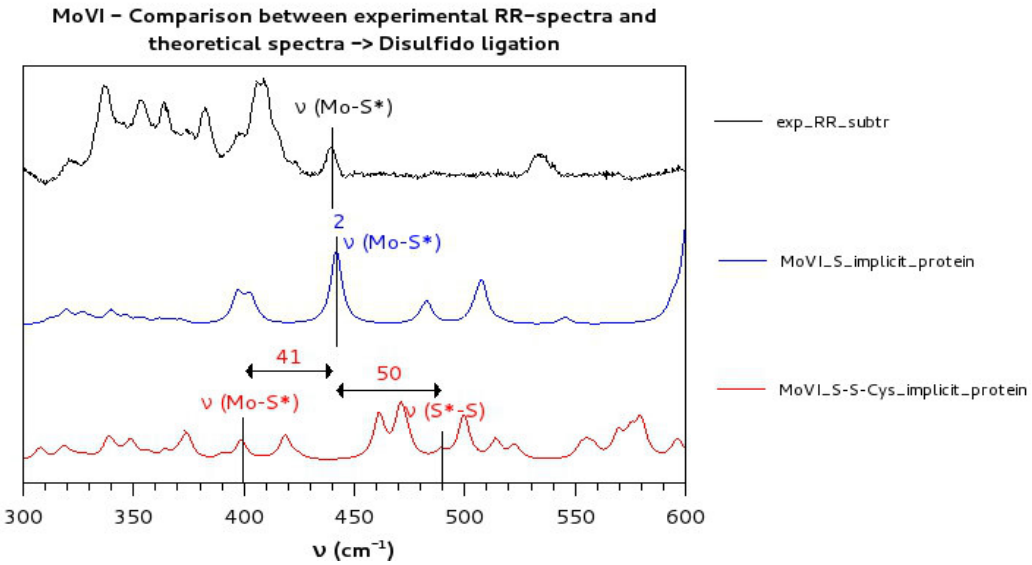

Figure 69. Calculated Raman-spectra (ε=4) for MoVI model 1 compounds: hexacoordinated with S*-ligand and Cys (MoVI_S; blue); and pentacoordinated with explicit treatment of disulfide bond between sulfido and Cys-ligand (MoVI_S*-S; red) and experimental subtraction RR-spectrum of WT (black) (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

In Figure 68 it can be seen that the MoVI_S* model yields a very good agreement to the measured data with the Mo-S* stretching signal at 441 cm^{-1} only shifted about 2 cm^{-1} from the experimental one at 439 cm^{-1} (as discussed in section 12.2.1.). The MoVI model 1 compound with included disulfide bridge yielded a Mo-S* stretching at 398 cm^{-1} which is downshifted from the experimental one by $\Delta\nu$=41 cm^{-1}. Furthermore, the S*-S vibration occurs at 489 cm^{-1} which is upshifted compared to the exp. band at 439 cm^{-1} by $\Delta\nu$= 50 cm^{-1} and shifted about 48 cm^{-1} from the Mo-S* stretching in the hexacoordinated MoVI_S* model. Therefore, the computed spectra is in less good agreement to the experiment than the MoVI_S* model. This makes the possibility of a sulfur-shift mechanism in the MoVI-state of the Moco less probable. Additionally, it further underlines that the band at 439 cm^{-1} in the RR-spectrum of the oxidized WT form of *Rc*FDH can be assigned to the Mo-S* stretching of a hexacoordinated Moco and that the cysteine-residue is directly ligate to MoVI. Calculations of models in vacuum yielded similar results (s. Appendix B).

13.3. Reduced MoIV-form

In section 12.3. it has been shown that pentacoordinated model 1 compounds with MoIV and not present Cys yield computed vibrational frequencies which are in good agreement to experimental ones. Also, Hartmann et al. [49] and Schrapers et al. [50], as well as Raajmaker et al. [52] proposed a dissociation of Cys upon formate-reduction of molybdenum. Therefore, the possibility of of an unbinding of Cys *via* a sulfur-shift mechanism, as well as possible positions and protonation states of Cys in the MoIV-form of the Moco have been investigated in more detail. Enthalpy-calculations for bond-breaking of the Cys-MoIV bond and a possible MoIV-OCHO bond will be discussed in section 13.5.

13.3.1. Cys-ligation

To determine the position of Cys at the MoIV-site, three additional model 1 compounds have been built: In the first two, cysteine is present, but remains unbound, either in its deprotonated (MoIV_Cys$^-$) or protonated form (MoIV_S*_CysH). In both models Cys is not ligated to the metal and placed approximately 5 Å away from MoIV. Thus, both models represent a pentacoordinated MoIV-center. The third model includes a direct ligation of Cys to MoIV (MoIV_S*_bound_Cys$^-$) and is thus hexacoordinated. All three models include a S*-ligand at MoIV. The models have been geometrically optimized and Raman-spectra have been computed subsequently, comparing them to the MoIV_S* model 1 compound as established in section 12.3. and experimental data. Here, two different RR-spectra of the reduced *Rc*FDH, measured under different conditions, have been used which differ only in small parts with conserved aignal at 439 cm^{-1} (s. Figure 70 for more details about reaction conditions). The results are shown in Table 19 and in Figure 70.

	ν (Mo-S*)(cm^{-1})	**ν (S*-S) (cm^{-1})**	**ν (Mo-Cys)(cm^{-1})**
Exp. Red. RR	439	?	?
MoIV_S* (model 1 + ε=4)	461	-	-
MoIV_S*_Cys$^-$ (model 1 + ε=4)	359/369	359/369	339/359/369
MoIV_S*_CysH (model 1 + ε=4)	448	-	-
MoIV_S*_bound_Cys$^-$ (model 1 + ε=4)	405	405	186/229-279

Table 19. Calculated vibrational frequencies for MoIV_S* model 1 compounds **(**ε=4**)** with different ligation sites of Cys-ligand (pentacoordinated Mo and Cys unbound to MoIV, but either charged (MoIV_S*_Cys$^-$) or protonated (MoIV_S*_CysH) or Cys directly ligated to MoIV (hexacoordnated Mo; MoIV_S*_bound_Cys$^-$)) and comparison to corresponding experimental RR-spectrum (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 10 mM formate; provided by Stefan Wahlefeld [73])

In the model with unbound Cys$^-$, the Mo-S* stretching occurs at 369 cm^{-1} (Δν=70 cm^{-1} away from the experimental band at 439 cm^{-1}), while in the model with unbound CysH it appears at 456 cm^{-1} (Δν=9 cm^{-1} shifted from exp.). The model with bound Cys$^-$ gives a Mo-S* and a S*-S stretching at the same frequency of 405 cm^{-1} (difference to the band at 439 cm^{-1} in the RR-data is about Δν=34 cm^{-1}). The latter band is in very good agreement to the experimental signal at ~405 cm^{-1}. The model with unbound Cys$^-$ is in bad agreement to the experiment. Actually no prominent band is visible in the region between 370 cm^{-1} and ~500 cm^{-1}. The best agreement to the RR-data (regarding the signal at 439 cm^{-1}) has been achieved for the CysH model (Mo-S* stretching at 448 cm^{-1}; upshifted from exp. band at 439 cm^{-1} about Δν=9 cm^{-1}) which is even better than for the MoIV_S* model without any Cys (here it the Mo-S* stretching is shifted about Δν=22 cm^{-1} from the exp. band at 439 cm^{-1}).

The good agreement of the three models (with CysH, with bound Cys and the one without any Cys present in the calculations) might indicate that in the reduced form of the enzyme/Moco we have three different possible ligation-situations: (1) The Cys is dislocated from the metal and either very far away (stabilized by Arg587 [49]; most probably right after possible disconnection from Mo) or (2) it is near the Moco and protonated (this might happen after possible H-transfer from either

$HCOO^-$ or the SH-ligand at MoIV [49],[54], see sections 13.6. and 13.7.). Since the model with bound Cys^- gives also a signal for the Mo-S* mode, which is near to the exp. one at ~405 cm^{-1}, (3) it can also be assumed that after the conversion from $HCOO^-$ to CO_2, Cys^- is returning to the MoIV center. This might be happening only at the end of the cycle (right before MoIV is back-oxidized to MoVI) and for a much shorter time-frame (than the other two ligation-situations).

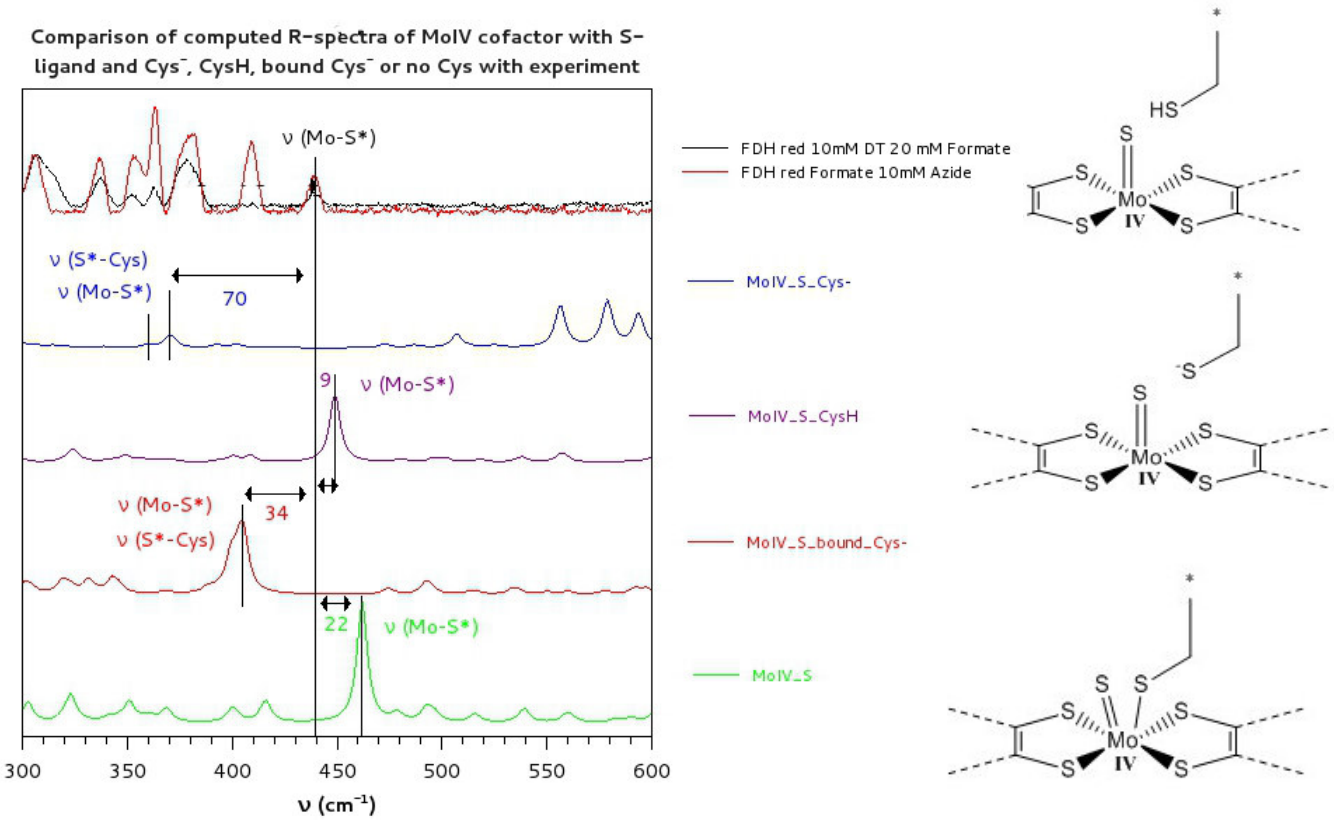

Figure 70. (left) Calculated Raman-spectra for MoIV_S* model 1 compounds (ε=4) with different ligation sites of Cys-ligand (pentacoordinated Mo; Cys not ligated to MoIV, but either charged (MoIV_S*_Cys^-; blue) or protonated (MoIV_S*_CysH; violet) or Cys completely absent (pentacoordinated MoIV_S*; green) and Cys bound (hexacoordinated Mo; MoIV_S*_bound_Cys^-; red)) and comparison to corresponding experimental data of reduced form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with either 10 mM formate (red) or 20 mM formate in 10 mM DT (black); provided by Stefan Wahlefeld [73]); (right) schematic representation of used model 1 compounds

These results are in contrast to the ones obtained for the MoVI-form of the Moco, in which the absence of Cys led to a significant decrease in quality of the computed Raman-spectra in comparison to the exp. (s. section 12.2.1.). In the case of MoIV, the models with completely absent Cys and a protonated and unbound cysteine yield the Raman-spectra with the best agreement to experimental data.

This further underlines the assumption that Cys is loosing its direct coordination to the metal (in contrast to the mechanism proposed by Maja et al. [54] and in agreement to the one from Hartmann et al. [49]), but only when molybdenum is in its +4 form.

Additionally, the modeling of a MoIV model 1 compound with unbound Cys^- and with a SH-ligand

instead of a sulfido-one, yielded interesting insights: During the geometry optimization procedure the proton of the SH-ligand immediately leaves the S*-atom and is transferred to the negatively charged sulfor of the Cys^--residue. This is a very important feature concerning mechanistic properties, since Maja et al. [54] proposed that a formal hydride from formate is translocated to the S*-ligand at the metal. It suggests that even if a hydride-transfer leads to the formation of a SH-ligand at the molybdenum, a proton is immediately transferred to Cys, from which it can be translocated deeper into the protein.

Conclusively, according to computational findings, Cys might be either negatively charged and far away from the Moco or protonated and in near vicinity of the Moco. Thus it looses its ligation to the metal in its +IV form. This leads to the suggestion (which will be discussed in more detail in sections 13.5. to 13.7.) that the sulfido-ligand might be accepting a formal hydride from formate, transmitting the two electrons from $HCOO^-$ to the metal. Afterward, Cys, which has left the coordination-site (good agreement of calculated Raman-spectra for both models with either unbound and protonated or absent Cys to the exp.), is further accepting the proton from the temporarily formed SH-ligand (formation of protonated cysteine; as shown by QM-optimizations). Thus a S*-ligand will be recreated (as determined to be present in the MoIV state and discussed in section 12.3.). At the end of the catalytic cycle Cys might be back-binding to the metal (agreement of computed and experimental frequencies around ~405 cm^{-1}). Thus, MoIV assumes different coordinations, according to its position in the reaction cylce. This will be discussed in greater detail in section 13.7. This process would be in agreement with observations from Hartmann et al. [49] and Maia et al. [54]. Because of the good agreement of several bands in different ligation-scenarios of MoIV to the experimental RR-spectra it can be assumed that the measured probes for RR might contain a mixture of different MoIV-states, with different ligations.

Observations of additional bands concerning S-S* stretching vibrations in the models with bound or unbound Cys^- to MoIV suggest that a sulfur-shift mechanism is probable in the MoIV-form (in contrast to the Moco with MoVI). This might be the mode of action for the dissociation of Cys from the metal and will be investigated in more detail in the next section.

13.3.2. Disulfide bridge formation

A MoIV model 1 compound with an included disulfide bridge has been prepared (s. Figure 67) and geometrically optimized. Subsequently computed Raman-spectrum has been compared to the pentacoordinated MoIV_S* model (with absent Cys) and experimental data [73]. The assumption has been made that the probe of the reduced *Rc*FDH and thus the corresponding RR-data might be composed of different species and may also include one with S*-S bond at the MoIV center.

The obtained results can be seen in Table 18 and in Figure 71. In contrast to the MoVI_S*-S models, the Raman-spectrum of the MoIV_S*-S model is in good agreement to the experiment as depicted in Figure 71. The Mo-S* vibration in the MoIV_S* model is about 22 cm^{-1} upshifted, compared to the signal at 439 cm^{-1} in the RR-spectrum. In the disulfide model, we still have two bands at 388 cm^{-1} (Mo-S* stretching) and 423 cm^{-1} (S*-S stretching),which are nearer to the measured band at 439 cm^{-1} ($\Delta\nu$=51 cm^{-1} and $\Delta\nu$=14 cm^{-1}, respectively) than has been observed for the spectrum of the MoVI_S*-S model. The S*-S stretching signal in the calculated spectrum is in good agreement to the experimental band.

Conclusively, the band at 439 cm^{-1} observed in all experimental RR-spectra of the reduced enzyme might not only be attributed to the Mo-S* stretching, but might also include contributions from a S*-S stretching motion. Therefore, a sulfur-shift is possible and seems likely to occur in the reaction

cycle. This is also a further indicator that the Cys-ligand is leaving the direct ligation sphere upon formate-reduction of Mo (as already discussed in section 13.3.1.) and that this happens *via* a sulfur-shift. The Cys-dissociation will be investigated further in section 13.5.

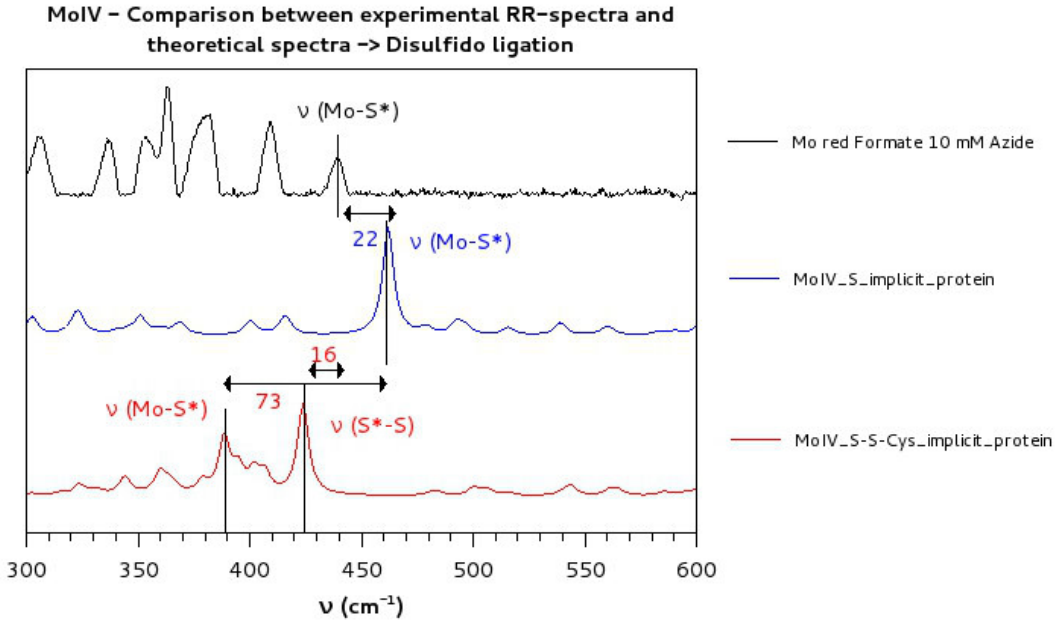

Figure 71. Calculated Raman-spectra for pentacoordinated MoIV model 1 compounds (ε=4) with S*-ligand and absent Cys-ligand (MoIV_S*; blue) or with explicit treatment of disulfide bond between sulfido and Cys-ligand (MoIV_S*-S; red) and comparison to experimental RR-spectrum of reduced form (black; 1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 10 mM formate; provided by Stefan Wahlefeld [73])

13.4. MoV-species

In previous EPR-measurements [52],[54],[56],[73],[196] the presence of a MoV-state in the catalytic cycle could be observed. This would imply a step-by-step electron uptake or release in contrast to the so-far predicted two-electron transfer processes. But it still remains questionable, if the observed signals in the corresponding EPR-spectra are due to a MoV-species or originate from an interaction between Moco and another species, possibly a counterion (azide or nitrate; used for stabilization of the enzyme during purification). Newest unpublished results [73],[196] suggest that even in the absence of the counterion the signal is still measurable which would underline the presence of MoV as possible reaction intermediate for molybdenum. Furthermore, in the EPR-measurements the influence of a proton in the second coordination sphere on the metal could be detected (e.g. [54],[73],[196]**)**. This may be indicate the formation of a SH-ligand at the molybdenum-site, as the sulfido-group gets protonated. Therefore, also the coordination sphere of a MoV-carrying Moco has been tested, especially the possible Cys-ligation to MoV.

MoV_SH models

Two different types of models with MoV have been built (s. Figure 72 and 73 on the right): One with S*-ligand and one with SH-ligand (since both may be present at MoV according to literature; e.g. [50],[52],[54],[56],[73],[196]). Cys is either deprotonated and near the MoV site (MoV_S(H)_Cys$^-$) or directly bound to MoV (MoV_S(H)_bound_Cys$^-$). Another possibility is that it is protonated and 5 Å away from the metal (MoV_S(H)_CysH). These models have been used for

QM geometry optimizations and subsequent Raman-spectra computations as described earlier. Since it is not possible to isolate the MoV-species for RR-measurements the computed Raman-spectra have been compared to the RR-spectrum of the reduced form of *Rc*FDH, because it has been proposed to include signals from a MoV-species. The results are shown in Figures 72 (left) and 73 (left) and Tables 20 and 21.

	ν (Mo-S*) (cm^{-1})	**Pterine-ring-deformations (cm^{-1})**	**δ (Mo-S*-H) (cm^{-1})**
Exp. Red. RR	439	?	?
MoV_SH_Cys⁻ (model 1 + ε=4)	365	~405	522
MoV_SH_CysH (model 1 + ε=4)	340	no signal near 400 cm^{-1}	528
MoV_SH_bound_Cys⁻ (model 1 + ε=4)	336	~405	639

Table 20. Calculated vibrational frequencies for MoV model 1 compounds (ε=4) and with SH-ligand and different ligation states of Cys: bound (MoV_SH_bound_Cys⁻) or unbound and either negatively charged (MoV_SH_Cys⁻) or protonated (MoV_SH_CysH); comparison with RR-spectra for reduced form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73])

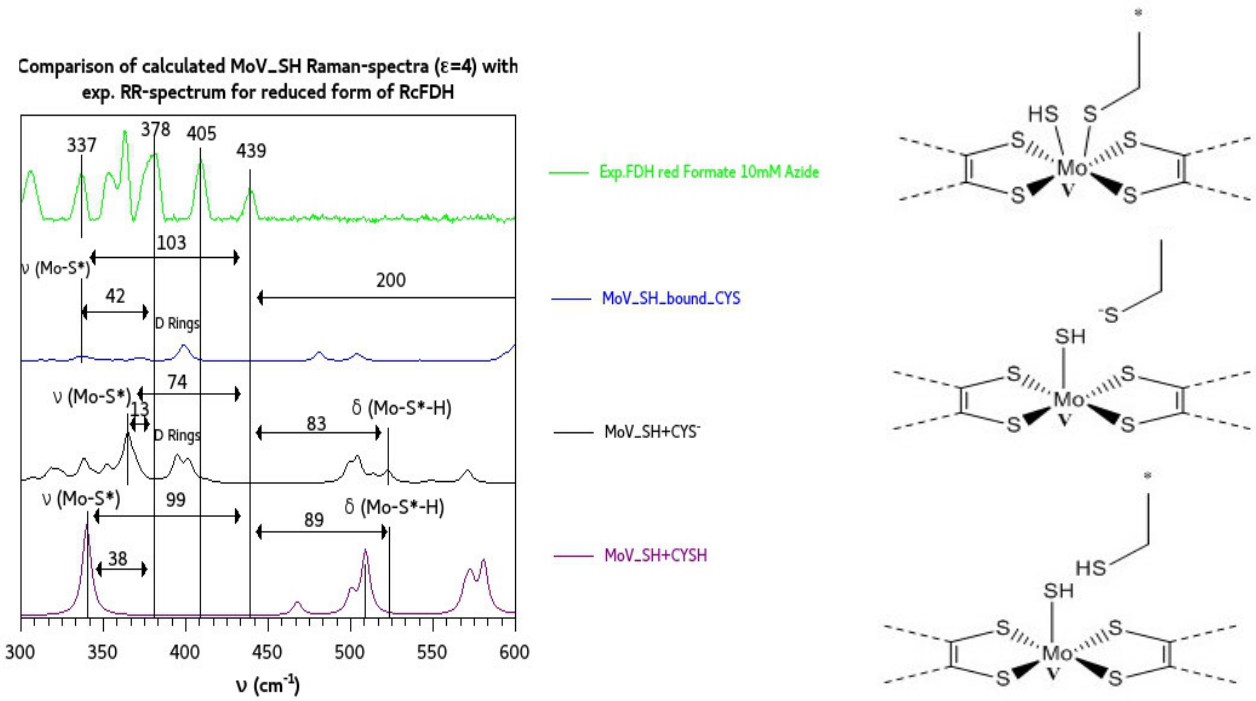

Figure 72. Calculated Raman-spectra for MoV model 1 compounds (ε=4) and with SH-ligand and different protonation states of Cys; not ligated to MoV and either negatively charged (MoV_SH_Cys⁻; blue) or protonated (MoV_SH_CysH; black); bound to MoV (MoV_SH_bound_Cys⁻; violet); comparison with RR-spectrum for reduced (green) form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73]); (right) schematic representation of calculated models

In the pentacoordinated model with unbound and negatively charged Cys^- the Mo-S* stretching signal appears at 365 cm^{-1} in near proximity to an unassigned band in the exp. RR-spectrum at 378 cm^{-1} ($\Delta\nu$=13 cm^{-1}). In both models with either bound (hexacoordination) or unbound but protonated Cys (pentacoordination) the Mo-S* stretching vibration can be seen at much lower wavenumbers at 336 cm^{-1} and 340 cm^{-1}, respectively. They are in good agreement to another unassigned band in the experiment at 337 cm^{-1}. But it should be noted that in comparison to each other, the Mo-S* vibration in the Raman-spectrum of the model with bound Cys shows only very small Raman-activity and is thus less comparable to the exp. In all three spectra, the Mo-S*-H angle bending vibration is shifted to higher wavenumbers at 522 cm^{-1} (unbound negatively charged Cys^-), 528 cm^{-1} (unbound protonated CysH) and 639 cm^{-1} (bound Cys^-) and are not in vicinity to any measured signals. In both Raman-spectra with bound and unbound charged Cys^- a signal at ~400 cm^{-1} can be seen which corresponds to ring deformations of the pterin moieties and are in near proximity to either the experimental band at ~405 cm^{-1} (see section 12.1.2.; Figure 58, right) or also to 378 cm^{-1} or the potential Mo-S* stretching at 439 cm^{-1} with shifts of $\Delta\nu$=~0-5 cm^{-1}, $\Delta\nu$=~22 cm^{-1} and $\Delta\nu$=~39 cm^{-1}, respectively. Thus the experimental bands at 405 cm^{-1} and 378 cm^{-1} might contain contributions from pterin ring deformations and MoV-S* stretching vibrations, respectively.

Due to the proximity of computed signals to the experimental bands at 405 cm^{-1}, 378 cm^{-1} and 439 cm^{-1}, the best model was the one with unbound and charged Cys^- (MoV_SH_Cys^-). This state might be a temporary intermediate in the catalytic cycle.

The model with unbound and protonated Cys (MoV_SH_CysH) is less probable to be representing any present state, since the appearance of both a SH-ligand and a protonated Cys as reaction intermediate is most unlikely (see sections 13.6. and 13.7. for more details): The proton at the sulfido-ligand may come from the formate (as suggested by Maja et al. [54]). In such a case, the Cys would still be deprotonated, since it is most likely the abstraction partner for the H^+ (as shown in section 13.3.1.). Therefore, the MoV_SH_CysH species can be ruled out.

The model with bound Cys (MoV_SH_bound_Cys^-) on the hand is still a possible intermediate, but less than the one with unbound and charged Cys^- (MoV_SH_Cys^-), due to less good agreement to exp. RR-data.

Conclusively, it may be that the reduction of MoVI to MoV might occur with a bound Cys which afterward is leaving the direct coordination site of the metal. In such a scenario a combination of both hexa- and pentacoordinated MoV-species would be likely measured in the reduced RR-spectra.

MoV_S models

Since it is not clear, if MoV carries a S*- or a SH-ligand, also computations with sulfido ligated to the metal have been carried out. Therefore, MoV_S* models (s. Figure 73, right) have been computed, both with unbound protonated Cys (MoV_S*_CysH) and bound Cys^- (MoV_S*_bound_Cys^-). A model with unbound and charged Cys^- is most unlikely, since after a possible proton-transfer from the SH-ligand to Cys (thus regenerating a S*-ligand), Cys would become protonated (as discussed in section 13.3.1. for MoIV with SH-ligation). This means that it is most unlikely that a reduced MoV exists with both S*-ligand and deprotonated Cys^- nearby. Therefore, this species has not been used for spectra calculations.

The results are shown in Table 21 and Figure 73. In the Raman-spectrum for the pentacoordinated MoV_S* model with unbound and protonated Cys (Mo_S*_CysH), the Mo-S* stretching signal is shifted (in comparison to the experimental band in the reduced RR-spectrum at 439 cm^{-1}) to higher

wavenumbers at 498 cm^{-1} by $\Delta\nu$=59 cm^{-1}. Therefore, it is in very poor agreement to the experiment regarding these two bands. But a signal, concerning the Mo-S (dithiolene) stretching motions at ~360 cm^{-1} is in good agreement to the exp. band at 378 cm^{-1} (shifted about $\Delta\nu$=18 cm^{-1}). The same goes for a band at 401 cm^{-1} for pterin-ring-vibrations which is in good agreement to the experimental band at 405 cm^{-1}.

In contrast to this, the Raman-spectrum for the hexacoordinated MoV_S* model with bound Cys (MoV_S*_bound_Cys^-) is in very good agreement to the experiment. Here, the signal for the Mo-S* stretching at 424 cm^{-1} is only shifted about $\Delta\nu$=15 cm^{-1} from the experimental band at 439 cm^{-1}. Furthermore, a second peak at 401 cm^{-1} (corresponding to deformations of the rings of the pterin moieties) is very near the experimental one at ~405 cm^{-1} ($\Delta\nu$=~4 cm^{-1}) and the one at 378 cm^{-1}.

	ν (Mo-S*) (cm^{-1})	**Pterine-ring-vibrations (cm^{-1})**
Exp. Red. RR	439	?
MoV_S*_CysH (model 1 + ε=4)	498	~410
MoV_S*_bound_Cys^- (model 1 + ε=4)	424	401

Table 21. Calculated vibrational frequencies for MoV model 1 compounds (ε=4) and with S*-ligand and different protonation states of Cys: unbound and protonated (MoV_S*_CysH) or bound (MoV_S*_bound_Cys^-); comparison with both RR-spectra for reduced or oxidized (WT) form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73])

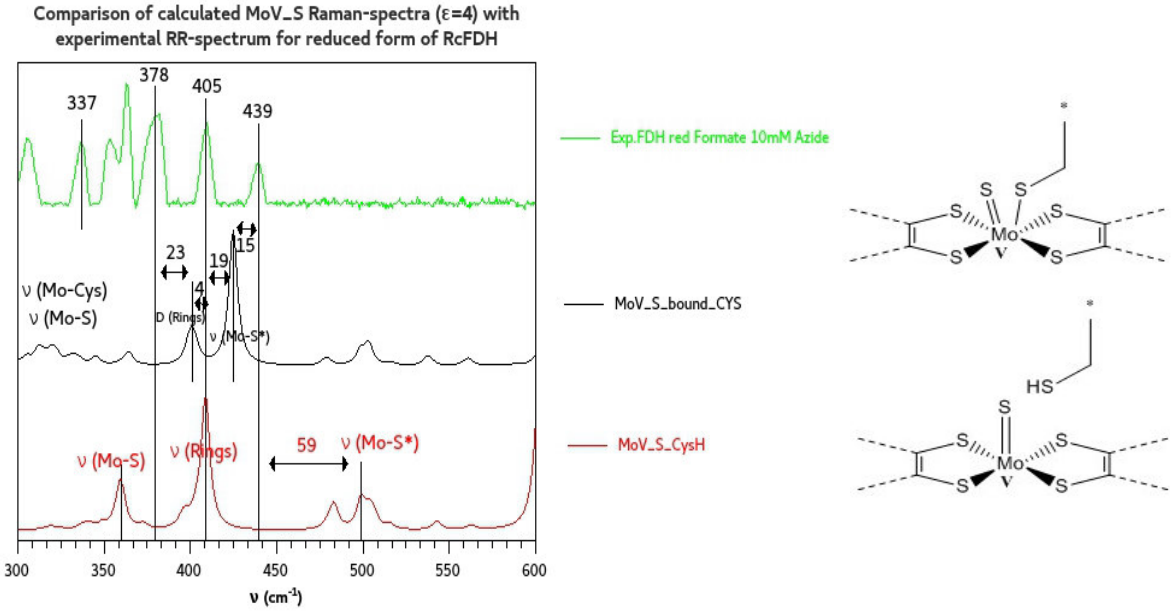

Figure 73. Calculated Raman-spectra for MoV model 1 compounds (ε=4) and with S*-ligand and different protonation states of Cys: not ligated to MoV and protonated (MoV_S*_CysH; black) or bound to MoV (MoV_S*_bound_Cys^-; violet); comparison with RR-spectrum for reduced (green) form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73]); (right) schematic representation of calculated models

Still, the MoV_S*_bound_Cys$^-$ model with hexacoordination represents the experimental RR-data in the best way, since two computed signals are in very good agreement to the measurements: Thus the bands at 378 cm^{-1} and 405 cm^{-1} in the reduced RR-spectrum might contain contributions from ring-vibrations of a Moco with MoV-S*-ligation and with bound Cys$^-$, whereas the signal at 439 cm^{-1} might also contain contributions from the MoV-S* stretching vibration. But, since also the pentacoordinated model shows a band near 405 cm^{-1} it might be that a mixture of both species is present, Whereas, the pentacoordinated model is of shorter lifetime and less probable, as can will be discussed in sections 13.6. and 13.7.

Conclusively, it can be assumed that the band at 439 cm^{-1} in the RR-spectrum of the reduced form of the Moco is also attributed to the MoV-S* stretching of a hexacoordinated MoV_S*-species with bound Cys, while the band at 378 cm^{-1} can be due to the MoV-S* vibration in MoV_SH complexes with protonated Cys or to ring deformations of the pterin groups in the MoV_S* model with bound Cys.

Both, the model 1 compounds MoV_SH_Cys$^-$ and MoV_S*_bound_Cys$^-$ are in very good agreement to the experimental RR-spectrum of the reduced *Rc*FDH and may thus be present in the catalytic cycle. It is possible that several MoV-species occur at different points in the reaction cycle (either in the reduction process MoVI → MoIV or in the backoxidation process of the metal MoIV → MoVI). As already discussed for MoIV-states, a mixture of different MoV- and MoIV-species might be present in the measured RR-probes This is discussed in further detail in section 13.7.

Future RR-experiments should target the signals around 500-550 cm^{-1}, which can be found for the MoV_SH models, because they could be excellent marker bands for further determination of possible MoV-intermediates. Due to the here shown results (in addition to EPR-results [54],[73], [196]), MoV-contributions to the catalytic oxidation of formate in *Rc*FDH should be accounted for in proposed mechanisms.

13.5. Bond-breaking enthalpies of MoVI- and MoIV-specimen

To finally decide, whether Cys is ligated to MoVI or MoIV, thermodynamic bond-breaking enthalpies have been calculated, concerning the binding of the Cys residue, as well as a formate molecule to a hexacoordinated MoVI- and MoIV-center, respectively. Therefore, two different model 1 compounds have been built and geometrically optimized: One with a bound Cys$^-$ at the Mo-center (MoVI/IV_S*_bound_Cys$^-$) and another with bound HCOO$^-$ at the molybdenum (MoVI/IV_S*_bound_HCOO$^-$); both with coordinated S*-ligand (1). Subsequently, the pentacoordinated MoVI/IV_S* systems without bound Cys$^-$ or HCOO$^-$ (2) have been geometrically optimized, as well as the ligand (Cys$^-$ or HCOO$^-$) by itself (3). The models are shown in Figure 74.

Figure 74. Schematic representation of calculated MoVI and MoIV model 1 compounds with S*-ligand present and: (left) MoVI/IV_S*_bound_Cys$^-$ (bound Cys); (right) MoVI/IV_S*_bound_HCOO$^-$ (bound formate)

ε=4	$\Delta H_{bond\text{-}breaking}$ (kcal/mol)
MoVI_S*_bound_Cys$^-$ (model 1 + ε=4)	+27.839
MoVI_S*_bound_HCOO (model 1 + ε=4)	+4.387
MoIV_S*_bound_Cys (model 1 + ε=4)	-26.843
MoIV_S*_bound_HCOO (model 1 + ε=4)	-5.670

Table 22. Calculated bond-breaking enthalpies for either Cys- or formate-ligation at both hexacoordinated MoVI- and MoIV-sites (ε=4)

Bond-breaking enthalpies have been generated by subtraction of enthalpy-values for (1) from the sum of the corresponding ones for (2) and (3), according to $\Delta H_{bond\text{-}breaking}$ = (2) + (3) – (1). The results are shown in Table 22 (s. Appendices D and E for additional material, like computations in vacuum, as well as differently composed models for enthalpy-computations).

In the oxidized form with MoVI, the dissociation of the MoVI-HCOO bond needs only +4.387 kcal/mol, while the MoVI-Cys bond is more stable, because +27.839 kcal/mol are needed to break this bond. For the Moco with MoIV it is *vice versa*: Here the formation of a hexacoordinated metal is less favorable, since both ligands would yield energy upon leaving the metal. Breaking of Mo-Cys bond is more feasible (-26.843 kcal/mol are gained). Dissociation of the one between MoIV and formate yields -5.670 kcal/mol. Therefore, MoIV exists most likely in a pentacoordinated form (as already suggested in section 13.3.1.). The differences in enthalpy between coordinated Cys^- and $HCOO^-$ to MoVI ($|\Delta(\Delta H_{bond\text{-}breaking})|$=23.452 kcal/mol) suggest that Cys is not leaving the coordination site in the oxidated Moco. Thus no free ligation position for formate is generated. The substrate might interact with the Moco *via* the sulfido-group and stays near the cofactor. Maybe by formal hydride-transfer and generation of a SH-ligand or by proton-transfer and seperate transfer of two electrons and in both the formation of a reduced metal in its +IV state. This is in agreement to the recent mechanism proposal of Maia et al. [54] who suggested that formate is not directly ligated to the metal-center. In the MoIV-form, the Cys-binding is lost, since here the Mo-Cys bond is energetically less stable ($|\Delta(\Delta H_{bond\text{-}breaking})|$=21.173 kcal/mol; as compared to the MoIV-OCHO bond). A free coordination position is thus created, with which the product CO_2 can possibly interact. This would be partly in agreement with the proposal from Hartmann et al. [49] and EXFAS-data which showed a MoIV-O bond to be present during the catalytic cycle [49],[50]. Also, this is in good agreement to the findings, presented in section 13.3.1. which suggested the dissociation of Cys from the reduced metal-site and possibly its return at the end of the cycle; see. 13.3.1. The return of Cys to the metal can be mediated by the temporarily formation of a S*-S disulfide-bridge with MoIV (as discussed in sections 13.3.2.).

These results are in agreement with conclusions from computations of models with different azide binding-sites (s. Appendix C for more detailed information). Azide, isoelectronic to formate, is an inhibitor for formate-oxidation in *Rc*FDH. DFT-computations of vibrational frequencies and subsequent comparison with IR-experiments [73] showed that azide is not directly binding to MoVI. Instead it stays near the Moco and might interact with the S*-ligand at the metal-site. MoVI remains hexacoordinated, with an all-sulfur coordination sphere. Furthermore, the energies obtained in vacuum shown in Appendix D and E display a similar trend as discussed here.

Contrasting this, computations of Leopoldini et al. [71] suggested the dissociation of Cys and direct ligation of formate to molybdenum. But these studies always assumed a starting structure with already bound formate to Mo. They did not investigate the probability of formate-ligation, but instead followed the suggestions of X-ray studies [52],[56]. Furthermore, Mota et al. [53] proposed also a direct ligation of formate, but these results have been revised later [54]. Therefore, these previous studies do not exclude the possibility of an unbound formate molecule.

In terms of completion it has to be mentioned that the differences in bond-breaking enthalpies between both MoVI-Cys and MoVI-OCHO bonds are small enough to be possibly compensated by rearrangements in the protein binding-pocket or energy-gains in other reaction-steps during the cycle. The same can be said about ligation of formate to the reduced MoIV state. Here, only 5.6 kcal/mole have to be compensated. Thus a direct coordination of formate to the molybdenum could not completely be excluded (s. Appendix F for further discussion). But due to the available computational data as shown here, it can be assumed that Cys remains coordinated to MoVI, even in the presence of formate and leaves the coordination-site only after reduction of the metal.

13.6. Proposed Mechanism for reversible formate oxidation at Moco-site

The results in chapter 12 and 13 lead to the assignment of different models and respective Raman-bands to the signals in the measured RR-spectra which are most probably corresponding to metal-ligand interactions. These results are presented in Table 23.

Enzyme-Species	Exp. frequencies (cm^{-1})	Assigned vibration (cm^{-1})	Model-assignment
Oxidized - WT	439	ν Mo-S* (441)	MoVI_S* (hexacoorindated; bound_Cys^-)
Oxidized - PreState	830/871	ν Mo-O* (874)	MoVI_O* (hexacoordinated; bound_Cys^-)
Reduced - WT	438/439	ν Mo-S* (461)	MoIV_S* (no Cys)
		ν Mo-S* (448)	MoIV_S*_CysH
		ν S*-S(Cys) (423)	MoIV_S*-S
	438/439	ν Mo-S* (424)	MoV_S*_bound_Cys^-
	~405	ν Mo-S* (405)	MoIV_S*_bound_Cys^-
	~405	ring deformation (401)	MoV_S*_bound_Cys^-
	378	ν Mo-S* (364)	MoIV_SH (no Cys)
	378	ν Mo-S* (365)	MoV_SH_Cys^-

Table 23. Assignment of possible vibrations to signals in the experimental RR-spectra of *Rc*FDH, furthermore, possible reaction intermediates are assigned (red attributes active oxidized MoVI-species, orange attributes for inactive desulfo MoVI-species, green attributes for reduced MoIV-species, blue attributes for reduced MoV-species); all models include coordination of molybdenum by two dithiolene moieties and exclusion of GDP-units (model 1 compounds), they have been geometrically optimized with DFT and usage of implicit protein environment (ε=4); all exp, data was contributed by Stefan Wahlefeld et al. [73] and were taken out as described in 12.1.2.

The two RR-spectra for both the active and inactive forms of the oxidized enzyme allow clear association with two models, the hexacoordinated models MoVI_S* and MoVI_O*, respectively, due to unique signals (439 cm^{-1} and 871 cm^{-1}, respectively).

In contrast to this, different models could be assigned to the RR-spectra of the reduced *Rc*FDH. This has consequences for the ligation-sphere and possible reaction intermediates during the catalytic cycle: While the oxidized enzyme has only two possible ligation-arrangements at the Moco, the reduced enzyme shows several: Pentacoordination with protonated Cys or completely absent Cys are possible, as well as hexacoordination with all-sulfur ligation. Also a SH-ligation, at least temporarily seems possible. Additional to MoIV-species, states with MoV are also possible. For all these models, corresponding computed vibrational frequencies can be assigned to experimental bands. This means that the Moco in the reduced enzyme undergoes changes during the reaction cycle, while the Moco in the oxidized form does not.

This leads to determination of the eight most probable reaction intermediates in the catalytic oxidation of formate, in the following named A to E, respectively. They are depicted in Figure 75 and in parts compared to experiments (contributions of Arg587 and His387 are according to Hartmann et al. [49] and were not tested here). EPR-results of Dufus et al. [73],[196] helped with MoV-contributions.

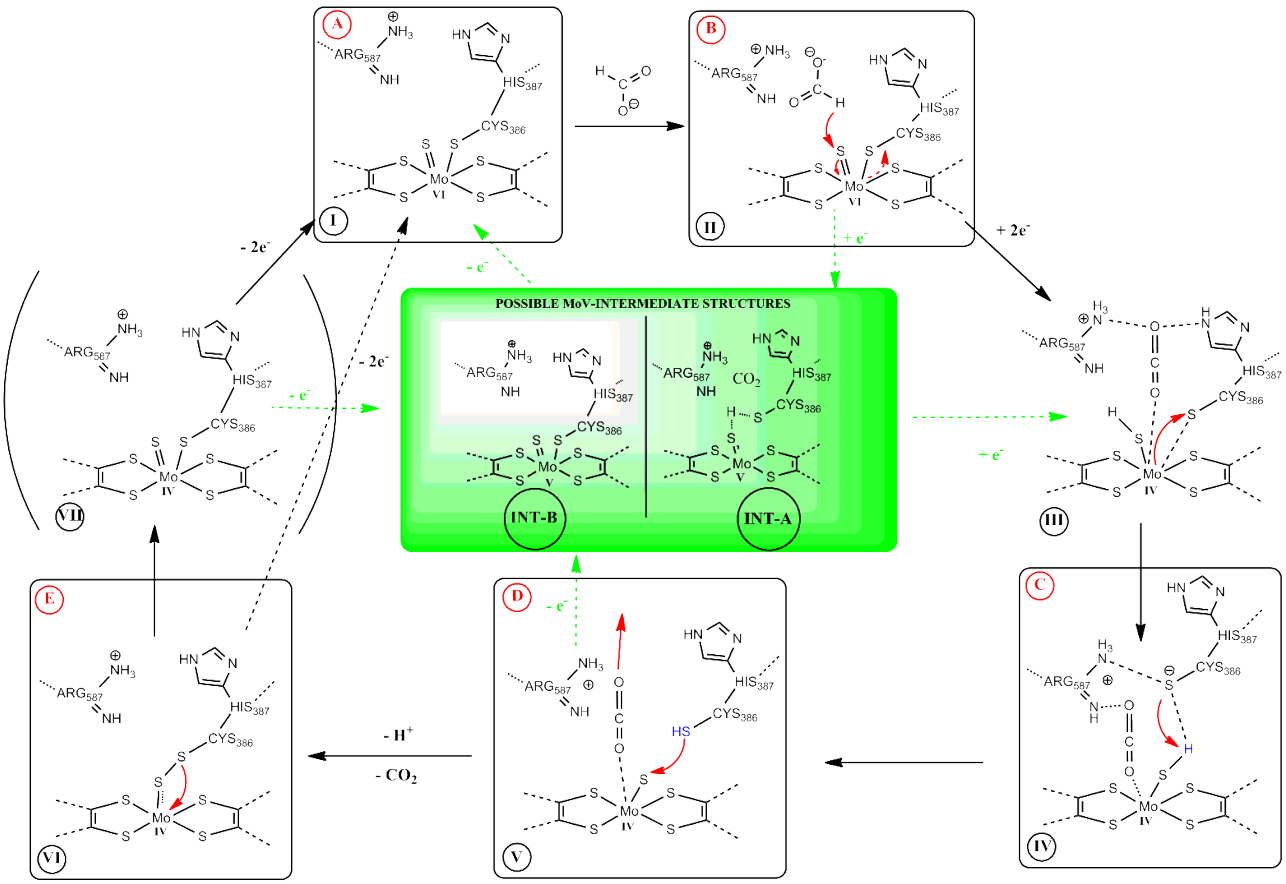

Figure 75. Proposed catalytic cycle for reversible formate oxidation at the Moco-site of *Rc*FDH using available experimental [49],[73],[196] and here presented computational data and includes formal hydride transfer from formate to sulfido-ligand, disclocation of Cys^- from the metal and sulfur-shift mechanistic parts, as well as possible MoV-intermediates

A. *Hexacoordinated MoVI, with all-sulfur coordination (two dithiolenes, sulfido-ligand, as well as native Cys-ligation).*

→ Computed Raman-spectra with all-sulfur hexacoordination of MoVI are in best agreement to experimental RR-spectrum of *Rc*FDH WT (s. 12.2.1). Absence of Cys-ligation led to worse agreement with experiment. Model with S*-ligand present at MoVI-site yield band at 441 cm^{-1} in near vicinity of experimental observed signal at 439 cm^{-1} in RR-spectrum of WT.

B. *Hexacoordinated MoVI, with native Cys-ligand and sulfido-group. The substrate-molecule (formate), is not directly ligated to MoVI, but stays in near vicnity; its negative charge may be stabilized by Arg587 [49]. The hydrogen may be interacting with the sulfido-ligand which may result ultimately in a formal hydride transfer from formate to the S*-ligand at MoVI.*

→ Computation of bond-breaking enthalpies revealed the more stable MoVI-Cys ligation, in contrast to the less stable MoVI-OCHO bond (s. 13.5.). $HCOO^-$ stays in near vicinity of MoVI, but does not exchange Cys as ligation-partner. Computations with inhibitor azide (s. Appendix C), revealed a similar behavior, with N_3^- not directly bond to the metal, but favoring a position near MoVI with Cys ligated to a hexacoordinated MoVI (the resulting vibrational frequencies are in good agreement to IR-experiments of Wahlefeld et al. [73]). Further, $HCOO^-$ might be interacting with S*-ligand, leading to the formation of a SH-ligand as also proposed by Maia et al. [54].

C./D. *Pentacoordinated MoIV-form, with short-living SH-ligand, whereas the proton of the thiol-group is immediately transferred to unbound and charged Cys, leading to a pentacoordinated MoIV-form with S*-ligand and nearby protonated Cys.*

→ Calculations of bond-breaking enthalpies revealed a less stable MoIV-Cys bond (s. 13.5.). Therefore, Cys is dislocated from MoIV, thus creating a vacant ligation-site at the metal. Possibly, CO_2 interacts with MoIV. Additionally, computed Raman-spectra with pentacoordinated MoIV and SH-metal-ligation (with absent Cys) showed in parts (signal at 364 cm^{-1}) agreement to the experimental RR-spectrum of the reduced enzyme (signal at 378 cm^{-1}). QM-geometry optimizations for models with SH-ligation at pentacoordinated MoIV and nearby unbound Cys^- showed fast proton-translocation from the thiol-group to Cys (s. 13.3.1.) which yields state D. Thus, species C (and therefore a SH-ligand) is a probable reaction intermediate, but only short-living and fast transformed into D. Finally, the Raman-spectrum of a model-compound with pentacoordinated MoIV, with S*-ligation and nearby unbound and protonated Cys yielded the best agreement (signal at 448 cm^{-1}) to experimental RR-spectra of the reduced species (signal at 439 cm^{-1}; s. 13.3.1).

E. *Pentacoordinated MoIV, with all-sulfur coordination, with S*-S disulfide-bridge formed between sulfido- and Cys-ligand.*

→ The formation of a disulfide-bridge between Cys and the sulfido-ligand at the MoIV-site could be observed during QM-geometry optimizations of models with MoIV, with S*-ligand and deprotonated and nearby Cys^- (s. 13.3.1). Additionally, the computed Raman-spectrum for MoIV model compound with explicit treatment of S*-S(Cys)-bound at the metal-site, showed a band for S*-S stretching (signal at 423 cm^{-1}) only 16 cm^{-1} displaced from the observed signal at 439 cm^{-1} in the exp. RR-spectra (s. 13.3.2). Therefore, this experimental band might also contain contributions of a disulfide-stretching motion.

VII. *Hexacoordinated MoIV, with all-sulfur coordination, with S*- and Cys-metal-ligation.*

→ Since the computation of the Raman-spectrum of the corresponding MoIV_S*_Cys$^-$ model yielded a band near 405 cm^{-1}, which is in very good agreement to an experimental signal at the same wavenumber, it is possible that even before the back-oxidation from MoIV to MoVI occurs, Cys is already bound to MoIV. This would be in contrast to the results of the bond-breaking enthalpies (as shown in 13.5.), maybe a rearrangement of the protein environment might help overcome energy-restrictions and make the back-binding of Cys to MoIV possible. This is not yet completely proven.

INT-A. *Pentacoordinated MoV, with all-sulfur coordination and thiol-ligation and deprotonated and charged Cys.*

and

INT-B. *Hexacoordinated MoV, with all-sulfur coordination and sulfido-ligand, as well as Cys-ligation.*

→ These two models showed good agreement to the experimental RR-spectra of the reduced enzyme (concerning bands at 378 cm^{-1}, 405 cm^{-1} and 439 cm^{-1}; see Table 23).

These eight states can be combined into a new mechanism for the catalytic oxidation of formate at the Moco-site of *Rc*FDH which combines the findings of Maia et al. [54] and Hartmann et al. [49] and other existing mechanistic/experimental studies (e.g. [50],[196]) and adds new features; see sections 2.3., 3.3. and 13.1. It is depicted in Figure 75:

Short explanation of mechanism, as depicted in Figure 75:

Starting with a hexacoordinated MoVI-species (**I**): Approaching formate is not directly ligating to the metal. Instead the hydrogen of the substrate is interacting with the S*-ligand, ultimately resulting in a formal hydride-transfer and the temporary formation of a SH-ligand (**II**). The reduction of MoVI to MoIV (**III**) leads to the dissociation of Cys from MoIV and the possible interaction of CO_2 with the metal (**IV**). Deprotonated Cys$^-$ abstracts a proton from the SH-group (**V**). The proton is further transferred in the protein matrix and Cys returns to the metal *via* disulfide-bridge formation (**VI**) (CO_2 is leaving the active-site). This can happen in two ways: Either Cys binds back to MoIV and then the metal releases two e$^-$ into the protein (from **VII** to **I**). Or first two electrons are translocated into the protein (MoIV → MoVI) and afterward Cys is returning to MoVÌ (directly from **VI** to **I**). The cycle can start again with a hexacoordinated MoVI-species. Also a step-by-step electron-transfer *via* MoV is possible by different routes. **II** can go to **IV** *via* **INT-A** and it is possible that the back-oxidation of the cofactor is happening by simultaneous back-binding of Cys and step-by-step electron-translocation (from **V** to **INT-B** to **I**). Or Cys is first returning to MoIV and then two electrons are transferred stepwise (from **VII** to **INT-B** to **I**).

13.7. Discussion of proposed mechanism and comparison to the literature

Main route of mechanism. The initial configuration **I** of the Moco is a hexacoordinate MoVI metal center, ligated by two dithiolene moieties, one Cys-residue and one sulfido-group. This has also been suggested by several findings in the literature: X-ray-studies of FDH-H of *E. coli* by Boyington et al. [56] found a Mo-S bond of about 2.7 Å which they associated with Mo-Cys-ligation. Similar results where discussed by Raajmakers et al. [52] who also included the possibility

of a sulfido-ligand due to a measured Mo-S* bond-length of 2.2 Å. According to these findings, Mota et al. [53] used also models with a all-sulfur hexacoordinated MoVI in their computations (based on X-ray of *Ec*FDH-H). Furthermore, both Maia et al. [54] (*Dd*FDH) and Raajmakers et al. [52] carried out the incubation of the enzyme with CN^- leading to an abolished catalytic activity. This inhibition, they suggested was due to the formation of an SCN^--species which demonstrated the presence and necessity of a sulfido-ligand to the reaction cycle. XAS-studies of *Rc*FDH by Schrapers et al. [50] and studies by Hartmann et al. [49] confirmed the hexacoordination of MoVI. In the former they found additional to the four dithiolene-suflur-Mo bonds, both a short (2.17 Å) Mo-S* bond and a long (2.6 Å) Mo-S bond. These they assigned to a sulfido- and a Cys-ligation, respectively. Robinson et al. [55] suggested a similar starting-coordination of MoVI of *Ec*FDH-H.

Starting with **I**, the presence of formate then initiates the catalytic process. In **II** formate is approaching MoVI, while stabilized by Arg587 and His387, as proposed by mutation-studies of *Rc*FDH of Hartmann et al. [49] which showed that the absence of these two residues led to diminished or completely abolished catalytic formate-oxidation activity. Formate is not directly coordinated to MoVI and Cys remains linked to the MoVI. This was also suggested by Maia et al. [54], since they did not found any evidence of the formation of a Mo-OCHO bond in their EPR-measurements. Furthermore, the hydrogen of formate may be interacting with the S*-ligand resulting in the temporary formation of a SH-ligand (as illustrated in **III**) and the transfer of two electrons, reducing MoVI to MoIV. This means a formal hydride is transferred from formate to Moco. Experimental results suggest also such a route: For example Porcher et al. [189] proposed a mechanism of H^+-reduction on a similar molybdenum-containing complex *via* direct formation of an OH-group from an initial O-ligand and e^--transfer. Furthermore, in EPR-studies the interaction between a MoV-species and a proton in the second coordination sphere could be observed [54],[73], [196]. Maia et al. found a hexacoordinated MoV-species with probable SH-ligation [54]. They discussed that the availability of the π-bond at S* makes it a good H-acceptor and its pka would change during the reversible MoVI ↔ MoIV redox-process [54]. Robinson et al. [55], following kinetic studies of inhibition of FDH with azide, proposed also the formation of a SH-ligand at molybdenum. Although, in contrast, they proposed direct ligation of substrate to metal (s. Appendix C). Finally, already early X-ray-studies by Raajmakers et al. [52] suggested the possibility of a SH-ligation and DFT-studies of Leopoldini et al. [71] based on *Ec*FDH-H took this into account accordingly.

The H^--transfer mechanism would explain, why a metal coordinated by an oxo-species (prestate) is inactive in contrast to the one with sulfido-ligation (active WT). The only differences between both ligands are the additional d-orbitals of the S* and the more negatively charged O* (higher electronegativity). The repulsive interaction between the H^- and the oxo-ligand, possibly explains its inability to accept the e^- from formate. In contrast to this S* is not as negative as O* and thus the interaction between S and H^- not as strongly repulsive and in addition the d-orbitals of S* might facilitate the binding of the formal hydride to the sulfido. This could be the reason, why no OH-ligand can be formed and the C-H bond of formate thus cannot be cleaved.

The next step is **III** to **IV**: In **III** MoIV is still hexacoordinated. Since it is energetically more favorable (s. section 13.5.), Cys leaves the coordination-site and the deprotonated Cys^- is stabilized by the Arg residue (**IV**), as has been proposed by studies of Hartmann et al. [49], in which mutation of this residue by a non-charged aminoacid led to a decrease in catalytic activity of the *Rc*FDH. Crystal-structure analysis of *Ec*FDH-H [52] found the displacement of a loop-region, containing Cys upon reduction of molybdenum. They found the distance between Cys and MoIV to be 12 Å [52]. This displacement of Cys from MoIV has also been found energetically favorable by DFT-

computations of Leopoldini et al. [71]**.** Mota et al. [53] demonstrated the strong repulsion between the sulfido-ligand and Cys (due to their negative charges) in their computational investigations of MoIV and MoV models. Iodoacetamid (IAA) assays of *Rc*FDH [49],[52] led to abolished activity of the reduced enzyme. This was interpreted such as that IAA reacts with Cys. This is only possible, if Cys has left its direct coordination to the metal. XAS-experiments [50] found the missing presence of a long MoIV-S bond which they attributed to the dissociation of Cys. And, furthermore, *Rc*FDH is able to process the reduction of NO_3^- which is only possible with direct ligation of NO_3^- to Mo and thus with a dissociated Cys from the Mo-center [49]. The latter of which was proven by catalytic inactivation of *Rc*FDH by IAA in the presence of NO_3^- [49].

In **IV**, following the SH-formation and dissociation of Cys from MoIV, Cys is now abstracting the proton from the SH-ligand, while CO_2 stays at the active site. This is in agreement with the proposed role of Cys by studies of Hartmann et al. [49]. The inactivity of *Rc*FDH [49],[52] while treated with IAA, suggested its direct involvement in the reaction cycle. Also a mutation of Cys to SER led to complete inactivity of *Rc*FDH [49]. Calculations [53],[71] showed also the H-transfer to the Cys. Although Hartmann et al. [49] and Mota et al. [53] suggested a direct abstraction of the hydrogen by Cys from formate and not from a SH-ligand, the function of Cys as proton-sink is the same, as presented here. Furthermore, it is possible that CO_2 interacts with the free coordination-site at MoIV, which would explain the findings of XAS-analysis of *Rc*FDH by Schrapers et al. [50] who detected a Mo-O signal (2.33 Å) in the reduced state of *Rc*FDH.

In **V**, the proton is already at Cys, leading back to a sulfido-ligation at an all-sulfur pentacoordinated MoIV. The possibility of pentacoordination and MoIV=S*-ligation has been discussed by Raajmakers et al. [52] and also shown by DFT-studies from Mota et al. [53]. XAS-results [50] revealed five short Mo-S bonds for a reduced RcFDH, interpreted as resulting from two dithiolene- and one sulfido-ligand. The proton at CysH is then translocated deeper into the protein matrix, while Cys returns to the molybdenum center. CO_2 leaves the active-site, since it is not directly bond to the metal and back-binding of Cys might change the protein environment in an unfavorable way for the molecule to stay.

In **VI** the return of Cys to MoIV *via* disulfide bridge-formation is illustrated. This is similar as in the proposed sulfur-shift mechanism by Mota et al. [53]. Their DFT-calculations suggested an activation of an inactive hexacoordinated MoVI-species by leaving of the Cys-residue *via* disulfide-bridge formation. Similarly, this active pentacoordinated Moco-form can be again inactivated by Cys ligating molybdenum via temporary formation of a disulfide-bond [53]. Also XAS-analysis by Schrapers et al. [50] showed an elongation of about 0.05 Å of the respective Mo-S* bond in the reduced probe which might be possible due to a S*-Cys-interaction.

Two different routes are possible at this point of the reaction cycle: From **VI** two electrons could be released and Cys returns directly to a MoVI-form resulting again in hexacoordinated form **I**. On the other side, it may be possible (due to agreements of computed Raman-spectrum of model with bound Cys to hexacoordinated MoIV with experimental RR-data [73]; as presented in section 13.3.1.) that an intermediate state **VII** is formed. Here, Cys is directly coordinated to the cofactor, while molybdenum is still in its reduced +4 form. Afterward, two electrons are released from the Moco (**VII**), leading back to the hexacoordinated form (**I**) with MoVI and the cycle can start anew. The existence of a hexacoordinated MoIV-state as in **VII** has been also suggested by EPR measurements of Maia et al. [54]: They found a MoV-state with six ligands attached to the metal, including Cys. Photolysis experiments showed no affection of the MoV-Cys bond. Thus they concluded that molybdenum in a reduced FDH has to exist in hexacoordination.

Mechanism including an EPR-active MoV-species. Additionally, it is possible that the electrons transfer not all at once, but step by step *via* a MoV-species. This was suggested by comparison of computed Raman-spectra with experimental RR-spectra of the reduced form of *Rc*FDH [73] (s. section 13.4.) and EPR-studies [52],[54],[56],[71],[73],[196] which found a MoV-species. Since it is not yet clear what ligands are present in the MoV-species (spectral comparison is not completely clear: SH- or S*-ligation; Cys still linked to MoV or not), several configurations are possible:

It is possible that, **II** undergoes a step by step e^--transfer with a short-living pentacoordinated MoV-intermediate **INT-A**, in which a SH-ligand and a dislocated and deprotonated Cys^- are present (see section 13.4.). Several studies discussed the presence of a pentacoordinated MoV-state with dislocated Cys (e.g. [50]). EPR-measurements [52],[54],[71],[73],[196] found a proton in the second coordination sphere of MoV which has been interpreted as the presence of a SH-ligand at MoV. Schrapers et al. [50], additionally, found an elongated Mo-S bond (of about 0.05 Å) in XAS-spectra of reduced *Rc*FDH which might be due to the formation of a SH-group. From **INT-A** the reaction then proceeds by another one-e^--transfer-step to yield the MoIV-form **IV**. Since the proton has been observed to be transferred to the nearby Cys^- very fast (s. section 13.3.1,), the corresponding MoIV/V-species (**IV/INT-A**) are supposedly only short-living.

The back oxidation of MoIV to MoVI can occur *via* two different MoV-mediated ways: It is possible that the e^--transfer and formation of MoV is occurring simultaneously with the returning of the Cys-ligand (from **V** to **INT-B** to **I**). MoIV would first be reduced to MoV, with a protonated Cys still nearby. Afterward, a proton is transferred from Cys further into the protein matrix and Cys binds back to MoV, leading to a hexacoordinated MoV-species, as discussed in 13.4. Thereafter, MoV is further reduced back to MoVI (**I**) and the cycle can start anew. And as a second route, it is also possible that the Cys-ligand is first returning (by S*-S-bridge formation) while Mo is still in its reduced +IV form, leading to **VII**. Afterward, the metal may be oxidized step-by-step *via* a MoV state (**INT-B**) and finally return to its initial state **I**. At this point of research, it is not possible to distinguish between these two ways. EPR-studies by Maia et al. [54], have determined a hexacoordinated MoV-state which is similar to **INT-B** (with the exception of a S*-ligation, insetad of SH-ligand, as proposed by Maia et al. [54]).

Since several studies are contrasting each other regarding the ligation-sphere of MoV and MoIV in the Moco (either hexacoordinated or pentacoordinated, with S* or SH) it might be that the measured probes of reduced FDH-species contain a mixture of different MoV/IV-species, with different ligations. According to the here presented computational and spectroscopic [73] results it is very plausible that MoV/IV assumes different ligation-scenarios at different points in the reaction cycle, as illustrated in Figure 75. This might explain the in parts controversial experimental results.

Finally it should be mentioned that recently a new mechanism has been proposed by Robinson et al. [55] also suggesting H^-/H^++$2e^-$-transfer to the sulfido-group, thus generating a SH-ligand at Mo. Although they also proposed direct formate-ligation to Mo (s. below), the similarities in their mechanism and the one presented here, shows the validity and importance of the here presented results concerning the formal hydride-transfer and the sulfido acting as H-acceptor.

Conclusively, the here proposed mechanism combines available experimental findings with new results from vibrational spectroscopy and computational studies. This leads to an expanded understanding of the catalytic cycle which overcomes restrictions of previous mechanisms.

Controversial areas in structural and mechanistic studies. It should be mentioned that inconsistencies between literature and the here presented results exist. These concern mainly three

subjects: (a) the ligands of MoVI, (b) the direct coordination of formate to molybdenum and (c) the likability of Cys to dissociate and the possibility of it accepting a proton from the Moco-site.

(a) MoVI-ligation-sphere. X-ray analysis of *Ec*FDH-H by Boyington et al. [56] suggested a hexacoordinated MoVI-state with two dithiolenes, a selenocysteine and controversially an OH-ligand. The latter they attributed to a 2.2 Å long bond at MoVI, they found in their studies. Furthermore, in reinterpretations of these studies by Raajmakers et al. [52], they were not able to distinguish between a S- or SH-ligation at a hexacoordinated MoVI. Additionally, they found a signal in EPR_studies which they interpreted as belonging to a SH-group at MoV. Following this observation, the DFT-models used by Leopoldini et al. [71] included a SH-ligand at the MoVI-site. Problematic with these findings is that e.g. the bond-length concerning the MoVI-X-ligation found in both crystal structures [52],[56] is better suitable for a Mo=S bond. Mo-SH bonds would be longer, and Mo-OH bonds shorter. Further, EPR-observation might indeed point to a MoV-SH scenario, but the SH-ligand might only exist in a reduced form of the enzyme/Moco. Finally, the structural analysis of Boyington [56] was subject to reinterpretations [52] of density-maps makeing it at least questionable in some areas.

(b) Direct coordination of formate. Some publications suggest the direct ligation of the substrate (formate) to molybdenum in the catalytic oxidation, although direct evidence is often missing. For example Boyington et al. and Raajmakers et al. [52],[56] suggested direct substrate-binding to Mo. The former found that NO_3^- ligated Mo which they concluded will be done similarly with $HCOO^-$. But although both structures share similarities, they are not the same and might exhibit different binding-strengths, when coordinating Mo. The latter [52] concluded from IAA experiments that $HCOO^-$ is occupying the vacant coordination-site at MoIV, since Cys dissociated. But dissociation of Cys does not necessarily evidence binding of formate. MoIV could stay pentacoordinated. Also IAA [52] worked only for already reduced enzyme, possibly meaning that only after reduction, does Cys dissociate (as also demostrated here by enthalpy-computations; s. 13.5.). In such a case the oxidation of formate to CO_2 has already occurred and therefore could not bind to molybdenum anymore.

DFT-studies by Leopoldini et al. [71] assumed formate-binding to Mo in their used models. Unfortunately, they only investigated the H-transfer from formate to Cys and not the possibility of formate-binding to the metal. Mota et al. [53] found in their DFT-computations a direct ligation of formate to Mo. The problem might be that their used methodologies included the manual choosing of different internal coordinates which then have been altered in order to find possible transition-states. Afterward, the corresponding educts and products for such transition-state have been calculated. Since the different internal coordinates have been chosen manually, the development of the system and thus the binding of formate to Mo might be artificially induced. It is not clear, if the system could not proceed in another way during the catalysis. Thus, both theoretical studies [53], [71] did not ultimately verify a direct coordination of formate to Mo.

Electrochemical experiments by Robinson et al. [55] revealed a strong selective binding-affinity of different inhibitors I (like N_3^-) to a MoVI-state. Inhibitors with strong net electron donation (EL) and high donor abilities were found to have the lowest dissociation-constants for Mo-I complexes. They concluded that since MoVI is e^--deficient, I most likely binds directly to MoVI. Formate, they suggested, is acting similar to the tested inhibitors. But. first of all, the tested inhibitors might facilitate different modes of action concerning the interaction with the Moco compared to formate, since they are chemically not equal. Second, electron-deficiency of MoVI might result in an "electron-poor" sulfido-group. This, in addition to its available d-orbitals makes the sulfido an excellent acceptor for a formal hydride. Furthermore, EPR-studies [54],[73],[196] suggest no direct

binding of N_3^- to MoV. See Appendix C for a more detailed discussion of inhibitor-binding studies. Robinson and coworkers [55] even suggested such an interaction between sulfido and I, but found it less probable. Finally, the selectivity of I to MoVI could also be based upon (as also mentioned by them [55]) protein arrangements, Cys-dissociation and MGD-reduction depending on the redox-state of Mo and is no direct proof of I (and thus $HCOO^-$) directly ligating molybdenum.

The only direct proof of a possible HCOO-ligation has been found by Schrapers and Hartmann et al. [49],[50]: In XAS-measurements of formate-reduced *Rc*FDH, they found the disappearance of a Mo-S bond and a newly formed Mo-O bond (2.23 Å). This they attributed to the dissociation of Cys and the formation of a formate-molybdenum-bond at the vacant position. Still, these findings remain questionable, since also a Mo-X signal could be measured which could not be assigned. Additionally, the Mo-O signal could also originate from a water-Mo interaction or from one between CO_2 and Mo, as suggested in this work. This makes their conclusion of a formate-molybdenum interaction via direct coordination at least questionable.

(c) Unprobable dissociation of Cys from molybdenum. Different studies suggest that Cys stays ligated to molybdenum during the whole catalytic cycle and is not able or insignificant for H-abstraction from formate: Boyington et al. [56] observed a Mo-Se (selenocysteine) bond in both oxidized and reduced enzyme. But reinterpretation of the crystal structure showed an unbound Cys in the reduced enzyme [52]. Furthermore, EPR-results [56] showed a proton in vicinity of Mo. Therefore, they suggested that a conserved His is abstracting the proton from formate [56]. But the signal in the EPR could also point to H being carried by the sulfido at MoV (e.g. [73],[196]). And mutation of His with Met in *Rc*FDH [49] yielded an enzyme which maintains catalytic activity. Thus, His seems not to be involved directly into the reaction cycle. Also even Boyington et al. [56] do not exclude the possibility that the H from formate is first taken by the sulfido-group, than transferred to Cys nearby and further on into the protein matrix. Maybe even His can temporarily act as proton-sink.

Maia et al. [54] found a hexacoordinated MoV-state in EPR-measurements and no dissociation of Cys from MoV during photolysis. Further they stated that IAA experiments did not work with the WT of FDHs. Therefore, they concluded that the reduced Moco, still carries a Mo-Cys-ligation, even in MoIV-form and it is less probable that Cys dissociates from Mo. Contradictory to this, a hexacoordinated MoV-species, does not exclude the existence of a pentacoordinated MoIV-state, with dissociated Cys. Also, as shown above, probably a mixture of different MoIV/V-states exist, depending on the reaction-step, resulting in different ligation-arrangements of MoIV/V. Furthermore, their statement concerning IAA experiments is incomplete, since with addition of NO_3^-, an inactivation of the WT could be seen [49]. Also, reduced probes were very sensitive to IAA-enabled inactivation [49],[52]. This shows the possibility of Cys to dissociate, at least with a stronger binding-partner like NO_3^- or at least in the reduced state of the Moco. Dissociation of Cys might depend on the redox-state of molybdenum. Also XAS-studies [50] showed the loss of the Mo-Cys bond upon reduction. Maia et al. [54] further suggested that Cys might generally not be able to abstract a proton from formate, since it is either too far away or bound to Mo. But, as discussed above (section 13.3.1. and 13.6., 13.7.), the proton from formate can first be abstracted by the sulfido-group and then further transferred to Cys,which seems to be a valid assumption.

Further investigations like e.g. improved RR-spectra, as well as additional QM-studies will further help to fill in more details in this mechanism: Since the appearance of different MoIV/V-states in the reaction cycle is still not completely clear, further theoretical evaluations have to be taken out. Including important protein residues in near vicinity of the cofactor into the computations (QM and QM/MM) might improve the obtained spectra and shed further light on the matter. Also, theoretical

simulation of EPR-spectra and the comparison of these to experiments may give further insight into the availability and the composition of MoV-species and their place in the catalytic process.

Additionally, the binding of $HCOO^-/CO_2$ to the active-site and its possible release has to be studied. The energy-costs to bind $HCOO^-$ directly to Mo in both +VI and +IV states might be small enough to be compensated by protein rearrangements or subsequent energy-gains in the reaction cycle. Therefore, a direct ligation of formate to molybdenum, as depicted in Figure 76 cannot be completely ruled out. The possible influence of this scenario on the catalytic cycle is discussed in more detail in Appendix F. Such an intermediate state INT (Figure 76) would be in agreement e.g. to findings of Leopoldini and Mota et al. [53],[71] and Schrapers and Hartmann et al. [49],[50], as well as Robinson et al. [55]. More ideas for investigations are discussed in chapter 14.

POSSIBLE INTERMEDIATE STRUCTURE

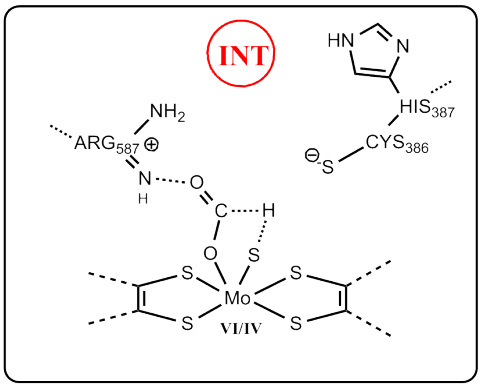

Figure 76. Possible intermediate INT with direct ligation of formate to molybdenum in its +VI or +IV state and H-abstraction from $HCOO^-$ by the sulfido-ligand at the metal

Finally, it has thus been shown that the here presented results are in agreement to publications. Contradictory findings in the literature are not as definitive, as demonstrated in the respective artciles, and in some cases can be even interpreted in such a way that they back the proposals given in this work. Therefore, the here proposed structures and mechanism (Figure 75) are a valid extension to the existing data. Missing information has been provided by density functional theory, as well as vibrational spectroscopy [73] and a more detailed and more complete understanding of the reaction cycle in the catalytic oxidation of formate in general and at the active site of *Rc*FDH in detail has been provided. This new mechanism combines findings from previous publications and adds new features. The results may be transferred to other Mo(co)-containing FDHs (enzymes) and will help future works in completely understanding the catalytic activity of FDHs and possibly in the creation of artificial enzymes with similar functionalities.

Chapter 14

Summary and conclusions of part V

CO_2-converting enzymes are of immense interest nowadays. They might provide a tool to reduce the amount of CO_2 in the atmosphere and simultaneously generate small carbon-based building blocks. These are important for all kinds of industries which until now depend mostly on fossil materials. Many molybdenum-containing enzymes are able to generate CO_2 out of small building-blocks and/or reverse the process. FDHs are one class of these enzymes and among others, the one from *Rhodobacter capsulatus* (*Rc*FDH) has been found to be able to catalyze both the oxidation of formate to CO_2 and the reverse process [48]. Therefore, it is of high interest to understand the mechanistic properties of it. Before understanding of the catalytic cycle can be achieved, the structure of the involved cofactors has to be clarified. In the case of FDHs the cofactors are conserved. At the active center the molybdenum-containing cofactor (Moco) is present. Several FeS-clusters are possibly responsible for electron transfer processes inside the enzyme. The formate-oxidation process is taking place at the Moco-site.

In this work, the focus has been on the study of the Moco-structure and the catalytic oxidation of formate. Therefore, several models have been derived, featuring different ligand-binding-scenarios based on the homology model (since no crystal structure had been available) as derived by Tillmann Utesch [49]. The models differed in the possible ligation of Cys (bond to Mo or not) and possible ligation-partners of molybdenum, such as S*, SH, O*, OH, S*-S(Cys) and $HCOO^-$. The used models excluded the GDP-units of the Moco, since they were not necessary for sufficient accuracy and allowed higher performance of the computations. The models have been geometrically optimized with QM-methods, using DFT- based techniques. Subsequently Raman-spectra have been computed and compared to experimental vibrational spectra [73].

The results verified a S*-ligand in the active and oxidized form of the Moco, while in the inactive (desulfo) form an O*-ligand has been found to be present. It has been further determined that MoVI is hexacoordinated, with two dithiolene moieties, one sulfido-group and the ligation of a native Cys. The reduced form of the Moco might exist in a mixture of states, with a pentacoordinated all-sulfur ligation sphere, with two dithiolenes and one sulfido-group ligated to MoIV, representing the most probable and stable form. Native cysteine has been shown to be able to assume different binding-situations in MoIV/V: In MoIV most probable is an unbound and protonated Cys. Furthermore, the coordination of a short-living SH-group to MoIV and MoV might be possible.

Following these findings, several follow-up calculations have been carried out, aiming at getting a better understanding of the mechanistic properties of the Moco. In the literature, currently two major reaction paradigms are discussed, concerning the formate conversion. One [49] is suggesting an unbinding of cysteine and a subsequent deprotonation of formate by this cysteine. In this process formate is directly ligated to the metal [49]. The other proposed mechanism [54] suggests that formate is not directly ligated to the molybdenum, but resides in near vicinity of the Moco and is stabilized by protein residues. A formal hydride is then abstracted by the sulfido-ligand at molybdenum [54],[72], resulting in the release of two electrons and CO_2. These two mechanisms have been taken into account, while deriving a new understanding of the catalytic process. The obtained results yielded a new mechanism:

Raman-spectra computations and comparison to experimental data [73], as well as computations of

bond-breaking enthalpies, showed the unbinding of cysteine from the metal only in its reduced +IV form. Thus, formate most probably does not ligate directly to MoVI, due to the highly stable MoVI-Cys bond. Furthermore, formate most likely interacts only indirectly with the cofactor *via* a formal hydride-transfer to the S*-ligand (but, as stated in the previous section: Changes in the protein environment or energy-gains in the reaction cycle, might help to overcome the potential energy-loss upon binding of formate to the metal). This leads to the reduction of MoVI to MoIV and the temporary formation of a SH-ligation. Since both Cys and $HCOO^-$ are energetically not favorable bound to MoIV, Cys is thus displaced from the metal. Afterward the SH-group transmits a proton further towards the unbound and deprotonated Cys. Subsequently CO_2 is released from the active-site and cysteine binds back to the metal, while the electrons are transferred deeper into the protein matrix. The on-and-off going of Cys may be done *via* a so-called sulfur-shift as reported also elsewhere [53]. The possible step-by-step electron transfer *via* a MoV-species has been addressed, although the results on this subject are not final. Nevertheless a MoV-intermediate is possible in the reaction cycle.

Although the proposed mechanism increases the understanding of the reaction cycle, not all parts are yet clear. There are several possible areas for future studies:

Possible deficiencies in the computed Raman-spectra might arise from shortcomings of the applied methods, such as missing scaling of the force-constants and thereby induced shifts in the vibrational frequencies. More sophisticated levels of theory (configuration-interaction (CI) and coupled-cluster (CC) methods), as well as, the influence of different exchange-correlation functionals and basis sets on the calculated spectra could be further investigated. These different methodologies might improve the results.

Additionally, the formal hydride transferred from formate to the S*-ligand at the MoVI should be further studied. This can be done with additional MoVI-models (sulfido-ligation) with either a H^+ or a H^- in near vicinity of MoVI. The addition of CO_2 or $HCOO^-$ to MoVI-models with S*-ligation may also provide insights into the matter. The same is true for the inclusion of Arg587 and His387 residues into the derived models. Geometry optimizations of these models, as well as subsequent Raman-spectra computations and comparison to experiments might also give more accurate frequencies. In addition, the possibility of formate binding directly to the molybdenum has to be further investigated. For this purpose, Raman-spectra calculation of models with S*- and SH-ligand in addition to a ligated formate at MoVI/IV and subsequent comparison to experimental data could be helpful. Furthermore, computations of bond-breaking enthalpies for the Mo-OCHO bonds with models including His387 and Arg587 should be helpful. A combination of such models with different DFT-functionals and maybe even HF-based electron-correlation methods might be shed further light on this subject. Finally, the application of transition state-scans, as well as Intrinsic Reaction Coordinate methods for finding reactants and products of these transition states could be of help in finding the correct arrangement of formate to Mo and possible hydride-transfer.

Further, the computation of EPR-spectra of the different MoV-binding situations and comparison to experimental data might help to readily determine the involvement of such an intermediate species in the reaction cycle.

Finally, the influence of the dithiolene and pterin parts of the cofactor has to be studied, since they are not redox-innocent and might contribute to electron-distribution and tuning of the redox-potential of the metal and its geometry (s. Appendix G). The here presented model-computations could be repeated with different oxidation-states of the pterins and further compared to experimental data.

Comparison between both O*- and S*-ligated models, as well as calculations of Mulliken charges and molecular electrostatic potentials might explain the reason of the inactivity of the desulfo-form. Electrostatic potential maps might shed light on differences in the charges in the Moco, depending on the respective ligand. Maybe distinctive differences between O*- and S*-ligation could be seen which might explain the reason of the inactivity of the prestate form. Addition of $HCOO^-$ (or simply the addition of a hydride in near vicinity of the Moco/sulfido-ligand) to these computations might explain the inactivation of the Moco with O*-ligation at MoVI. Also extensive analysis of the MO-structure of these two models have to be carried out. It may be possible that the oxo-ligation at MoVI leads to a change in the ligand-field and thus in differences in the electronic MO-structure of the metal-containing cofactor.

Such calculations would further enhance the strength and accuracy of the here presented mechanism for the catalytic oxidation of formate.

Conclusively, the here proposed structures and mechanism helps to get a better understanding of formate oxidation and may provide a useful starting-point for future studies. Furthermore, the here derived QM-models will be of use in subsequent investigations of Mo-cofactor-containing systems.

Final Remarks

Finally, it can be said that the usage of theoretical methods provide a useful tool for understanding biological matter and the underlying processes. Experiments alone are not able to deliver details on an atomic level. But this is often of utter importance, since mechanisms can mostly be understood on this scale. The combination of both experiments and theory yields the desired quantities.

Although biological systems are often consisting of thousands and thousands of atoms and thus limit the application of quantum chemical approaches, several theoretical methods can be applied to them. It is possible to simulate whole conformation spaces *via* classical MD-simulations. Further, hybrid QM/MM methods provide an opportunity to investigate interesting areas, like active centers, in more detail *via* quantum mechanics, while the influence of the rest of the system is treated on a molecular mechanics level. And it is also of great benefit to model only the cofactor directly by quantum mechanics. The additional calculations of spectroscopic data proved very helpful in delivering insight into dynamics and structure of the investigated systems. The continuing improvement of these methods by introduction of e.g. polarized charges, more complete structural models and so on, delivers a growing set of tools for investigating biological systems.

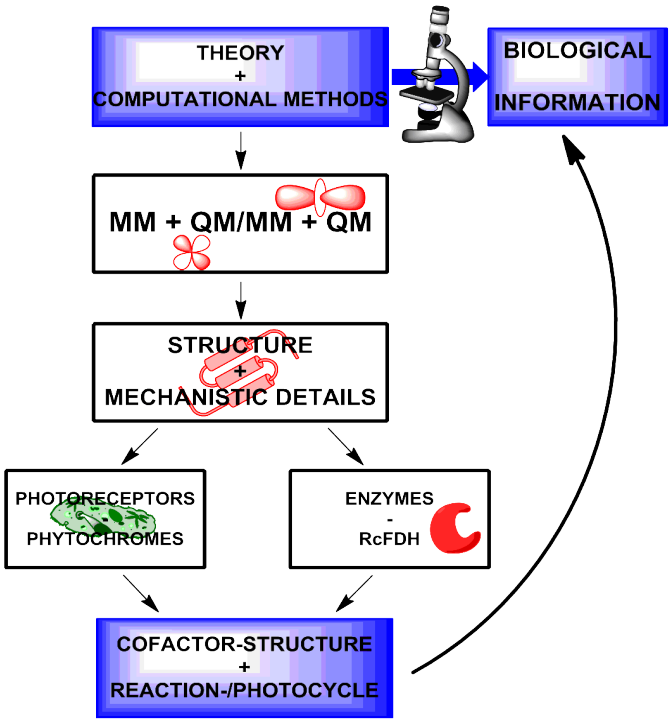

Figure 76. Evaluation of biological information *via* computational methods leads to the completion, prediction and explanation of experimental results concerning biomolecules

As shown in this work, all these methodologies have been successfully applied in studies of different proteins, such as photosensors, as well as catalytic active enzymes. Different methods, such as MM-MD, QM/MM and QM (and their improvement, such as the development and testing of a polarized force-field for phytochromes) have been used to determine the complete structure of cofactors (metal-dependent or photosensible) and to estimate mechanistic properties of the respective systems such as determination of important parts of the photocycle in phytochromes and key-steps in the catalytic oxidation of formate in FDHs. This way, computational chemistry, i.e. biomodeling, helped to complete experimental studies like filling-in of missing hydrogen-positions in X-ray structures, determination of ligands at the Mo-site in FDHs and explain exp. findings for

both phytochrome [26] and FDH-studies [73].

The computational studies, which have been taken out in this work, provided, furthermore, a new perspective for several questions, currently discussed in the scientific community concerning structure and mechanism of phytochromes, as well as formate dehydrogenases.

Last, but not least, all the described studies and their results, showed one thing above all – the beauty and perfection which is nature. For what we have to work hard to even understand, left alone to replicate or create ourselves, nature has found an often simple and most efficient way. Every question answered gives us a new perspective. And every new view we obtain leaves us with awe and wonder. The process of understanding nature and the universe humbles us and shows us our place in it – this is the reason I became a scientist and this is why I want to open the door wide and see right through it into the face of creation.

Acknowledgments

I want to thank first of all Prof. Maria Andrea Mroginski for the opportunity to work with her on these interesting topics, as well as for her guidance and support during my Phd-studies. My sincere acknowledgments go also to the TU Berlin for providing the possibility to work with Prof. Maria Andrea Mroginski.

Furthermore, I want to thank Prof. Silke Leimkühler for collaboration, as well as judging my thesis.

Sincere thanks to Prof. Hildebrandt for scientific discussions and fruitful collaborations.

My thanks to Dr. Patrick Scheerer, Dr. Stefan Wahlefeld, Dr. Ingo Zegber, Dr. Franzisco Escobar, as well as Maria Fernandez for provision of experimental data and many interesting discussions.

Jürgen Krauss I want to thank for his never ending help in IT matters.

For the computational resources I would like to thank the Deutsche Zentrum für Hochleistungs-rechnen HLRN, as well as the Institute of Mathematics at the TU Berlin.

This work has been financially supported by the SFB 1078 and UNICAT. Thanks go to you.

Thanks to all members of AG Hildebrandt and AG Mroginski for scientific help, guidance and last, but not least many hours of laughter, joy and above all, friendship.

Before I close this work I want to say thank you, to the people I value most and who made this journey possible:

I thank my parents, for bringing me into this fascinating world and teaching me to never stop asking and wondering, to give my best in everything and to once in a while just relax and take it easy. Their love give me strength.

Thanks to my friends and all the people I met on the road who were there for me when I needed them and who encouraged me and gave me their friendship and joy.

Most importantly, I want to thank my girlfriend Maria, for all she is giving me, since the first day we met: All the love you shelter me in and all the warmth and comfort you bring to me. You give me strength, when I need it most, you give me courage, when I am afraid and you show me wisdom, when I am lost. Your support helped me on this journey in countless situations. I love you!

Thanks to all of you for your help and support!

Appendix

A. Raman-spectra computations of model 1 and 3 compounds of Moco in *Rc*FDH in vacuum (ε=1) or protein environment (ε=4)

In Model 1 compounds (see Figure 61 for model-description) the complete GDP-units of the native Moco are ommitted. In model 3 compounds only the guanine units of the native Moco of *Rc*FDH are ommitted. The tables include values for the corresponding model 1 and 3 calculations (either with ε=1 or ε=4).

Oxidized form of Moco (MoVI)

	ν (Mo-S) (cm^{-1})	**ν (Mo-Cys) (cm^{-1})**	**δ (S-Mo-S) (cm^{-1})**	**δ (Mo-O-H) (cm^{-1})**	**δ (Mo-S-H) (cm^{-1})**	**ν (Mo-S*) (cm^{-1})**	**ν (Mo-O*) (cm^{-1})**
Exp. Subtr. RR (WT)	?	?	?	-	?	439	-
Exp. Subtr RR (desulfo-form)	?	?	?	?	-	-	830/871
MoVI__O* (model 1 + ε=4)	337/365	347	232/332	-	-	-	874
MoVI__O* (model 1 + ε=1)	336/358	344	233/344	-	-	-	889
MoIV_O * (model 3 + ε=1)	~330/352	338	~330	-	-	-	875
MoVI__OH (model 1 + ε=4)	377/383	~250	130/189	767/786	-	-	622
MoVI_OH (model 1 + ε=1)	336/378	336	190/383	770/787	-	-	630
MoIV_OH (model 3 + ε=1)	~330/363	-	201/~330	771/799	-	-	596
MoVI__S* (model 1 + ε=4)	332/353	330/346	111/211	-	-	441	-
MoVI_S* (model 1 + ε=1)	327-353	327-353	183/212	-	-	452	-
MoIV_S* (model 3 + ε=1)	325-360	318-335	292/335/340	-	-	432	-
MoVI__SH (model 1 + ε=4)	354/376	278/346	183	-	640	354	-
MoVI_SH (model 1 + ε=1)	365/386	278/362	181	-	640	386	-
MoIV_SH (model 3 + ε=1)	312-360	295/312/334	~192/~330	-	682	334	-

Table A1. Vibrational frequencies for marker bands of different hexacoordinated models with MoVI, differing in ligand (O*, OH, S* or SH), model-size (model 1: no GDP-units; model 3: no guanine-parts) and either ε=4 or ε=1, compared to experimental subtraction RR-spectra (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

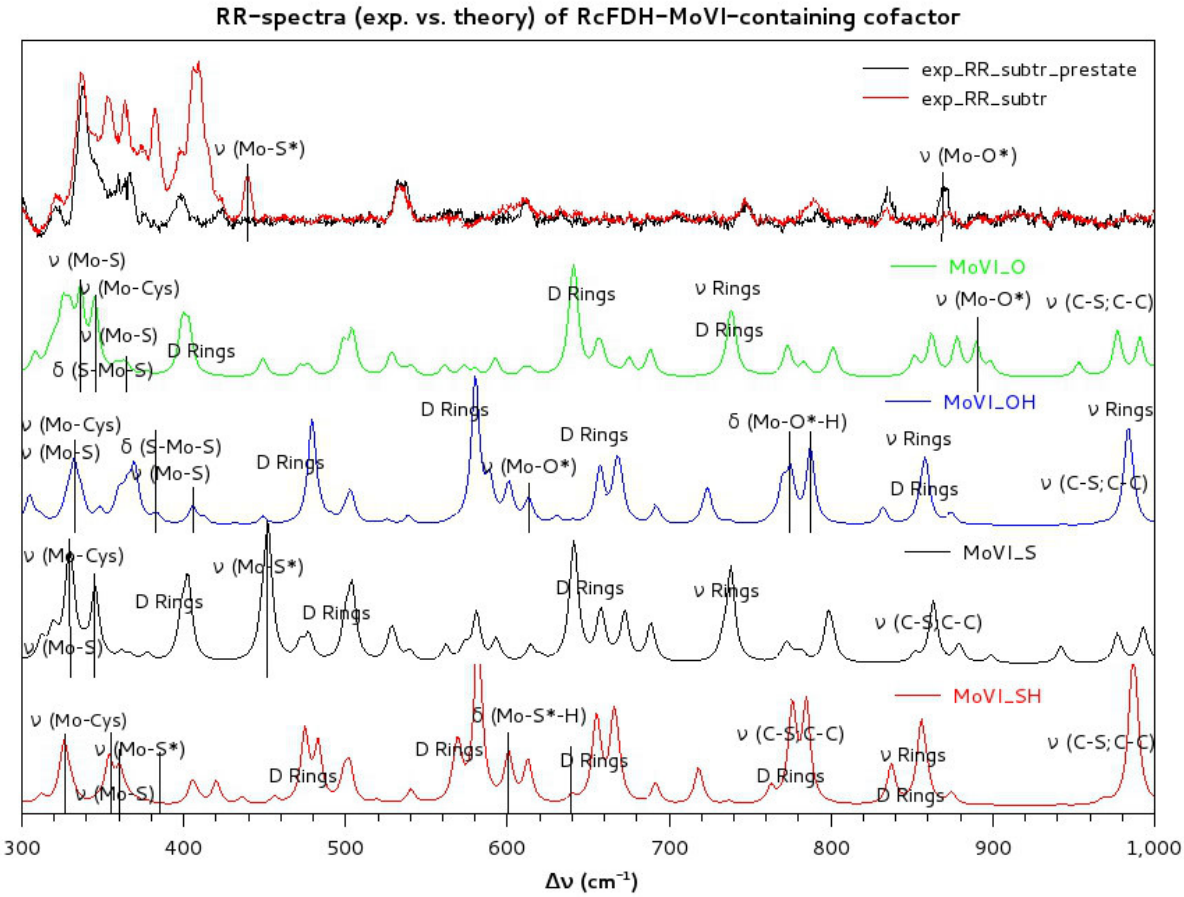

Figure A1. Calculated Raman-spectra of model 1 compounds (ε=1) for different ligand binding motifs at hexacoordinated MoVI (green: MoVI_O*, blue: MoVI_OH, black: MoVI_S*, red: MoVI_SH) and comparison with experimental subtraction RR-spectra: red: WT, black: desulfo-form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

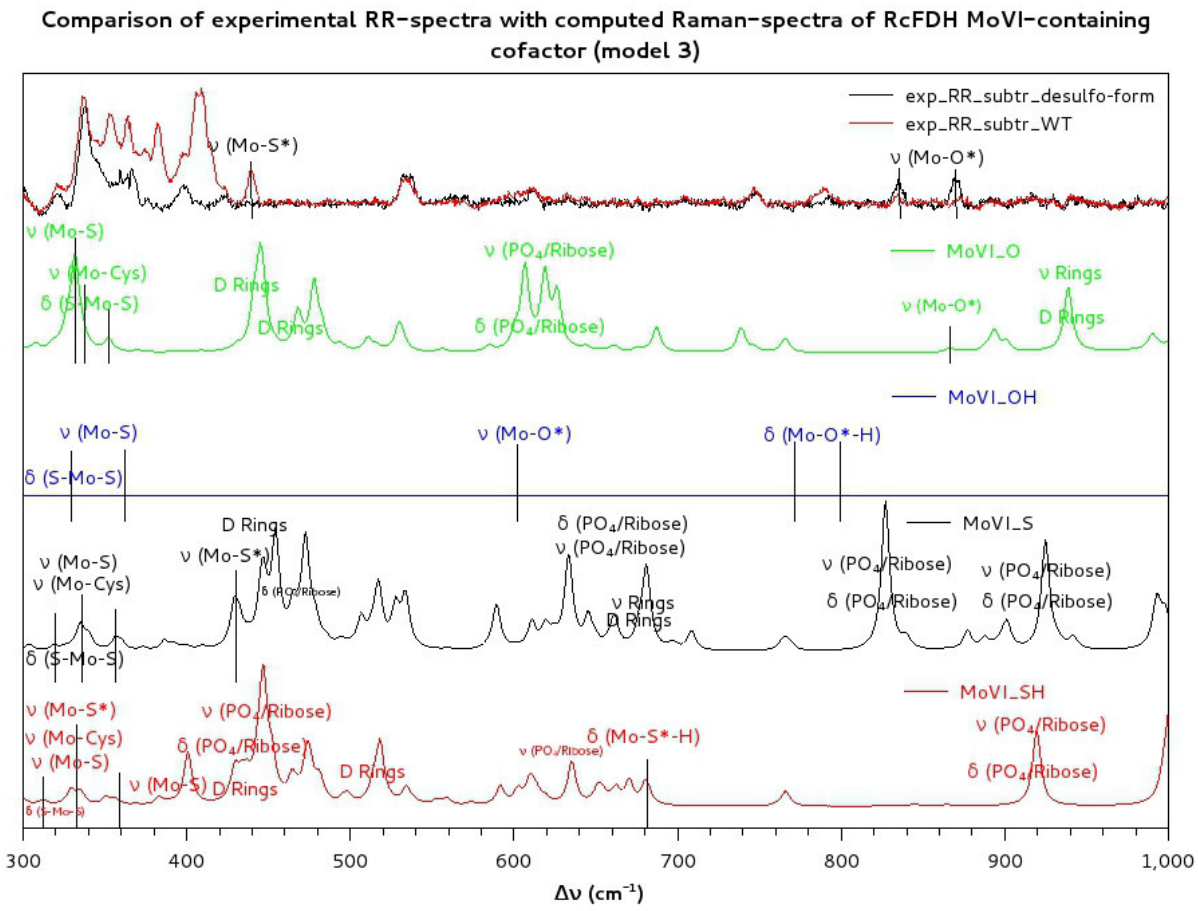

Figure A2. Calculated Raman-spectra of model 3 compounds (ε=1) for different ligand binding motifs at hexacoordinated MoVI (green: MoVI_O*, blue: MoVI_OH, black: MoVI_S*, red: MoVI_SH) and comparison with experimental subtraction RR-spectra: red: WT, black: desulfo-form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

NOTE: In Figure A2 the MoVI_OH spectrum is a blank line, since there was a problem in calculation of the polarizabilities and therefore it was not possible to gain any Raman activities.

In the Raman-spectrum of model 3 compound MoVI_SH a band at ~400 cm^{-1} can be seen which is due to a Mo-S (dithiolene) stretching. Although it is near the experimental signal at 439 cm^{-1} it is in worse agreement than the Mo-S* stretching in the MoVI_S* model 3 compound (s. Table A1). H/D-exchange experiments by Wahlefeld et al. [73] yielded no measurable shift of the corresponding band at 439 cm^{-1}, making it less possible that a Mo-SH-ligation may be present. In this work a MoVI_S* ligation-scenario has been assumed.

Also, in the MoVI_O* model 3 compound a signal at ~440 cm^{-1} can be seen which is attributed to ring-deformations of the pterin moieties. Since this band cannot be seen in spectra concerning model 1 compounds (either with ε=1 or ε=4) it is very unlikely that this band at ~440 cm^{-1} is reflecting the experimental one at 439 cm^{-1} in the active WT of *Rc*FDH.

Reduced form of Moco (MoIV)

	ν (Mo-S) (cm^{-1})	δ (S-Mo-S) (cm^{-1})	δ (Mo-O-H) (cm^{-1})	δ (Mo-S-H) (cm^{-1})	ν (Mo-S*) (cm^{-1})	ν (Mo-O*) (cm^{-1})
Exp. RR	?	?	?	?	439	?
MoVI_O* (model 1 + ε=4)	323-367	175	-	-	-	915
MoVI_O* (model 1 + ε=1)	342	364	-	-	-	938
MoIV_O* (model 3 + ε=1)	315/330-355	~170/~350	-	-	-	926
MoVI_OH (model 1 + ε=4))	342-381	178	680	-	-	640
MoVI_OH (model 1 + ε=1)	364	321	690	-	-	658
MoIV_OH (model 3 + ε=1)	343/365/387	156/~350	756	-	-	633/756
MoVI_S* (model 1 + ε=4)	322-360	182	-	-	461	-
MoVI_S* (model 1 + ε=1)	319/352	352/483	-	-	483	-
MoIV_S* (model 3 + ε=1)	309/328/360	~170/~350	-	-	467	-
MoVI_SH (model 1 + ε=4)	344-386	175	-	519	364	-
MoVI_SH (model 1 + ε=1)	363	-	-	541	541/368	-
MoIV_SH (model 3 + ε=1)	343-362	~170/~350	-	548	347/348	-

Table A2. Vibrational frequencies for marker bands of different pentacoordinated MoIV models, differing in ligand (O*, OH, S* or SH), different model sizes (either model 1 or model 3 compounds) and possible usage of implicit protein environment (either ε=1 or ε=4) and compared to experimental RR-spectrum (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73])

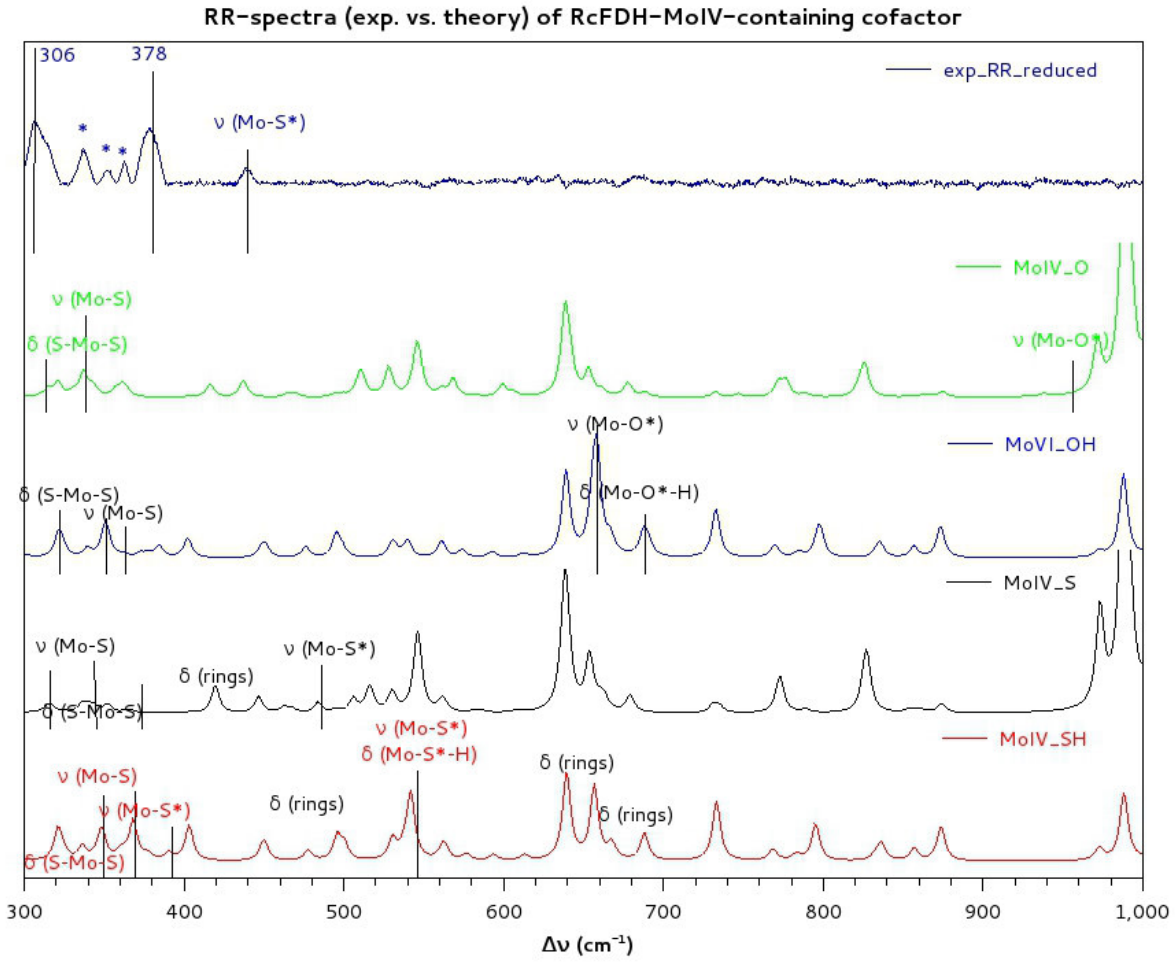

Figure A3. Calculated Raman-spectra of model 1 compounds (ε=1) for different ligand binding motifs at pentacoordinated MoIV (green: MoIV_O*, blue: MoIV_OH, black: MoIV_S*, red: MoIV_SH) without explicit treatment of Cys-ligand and comparison with experimental RR-spectrum (asterisk denotes potential FeS-cluster vibbrations): dark-blue: reduced form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73])

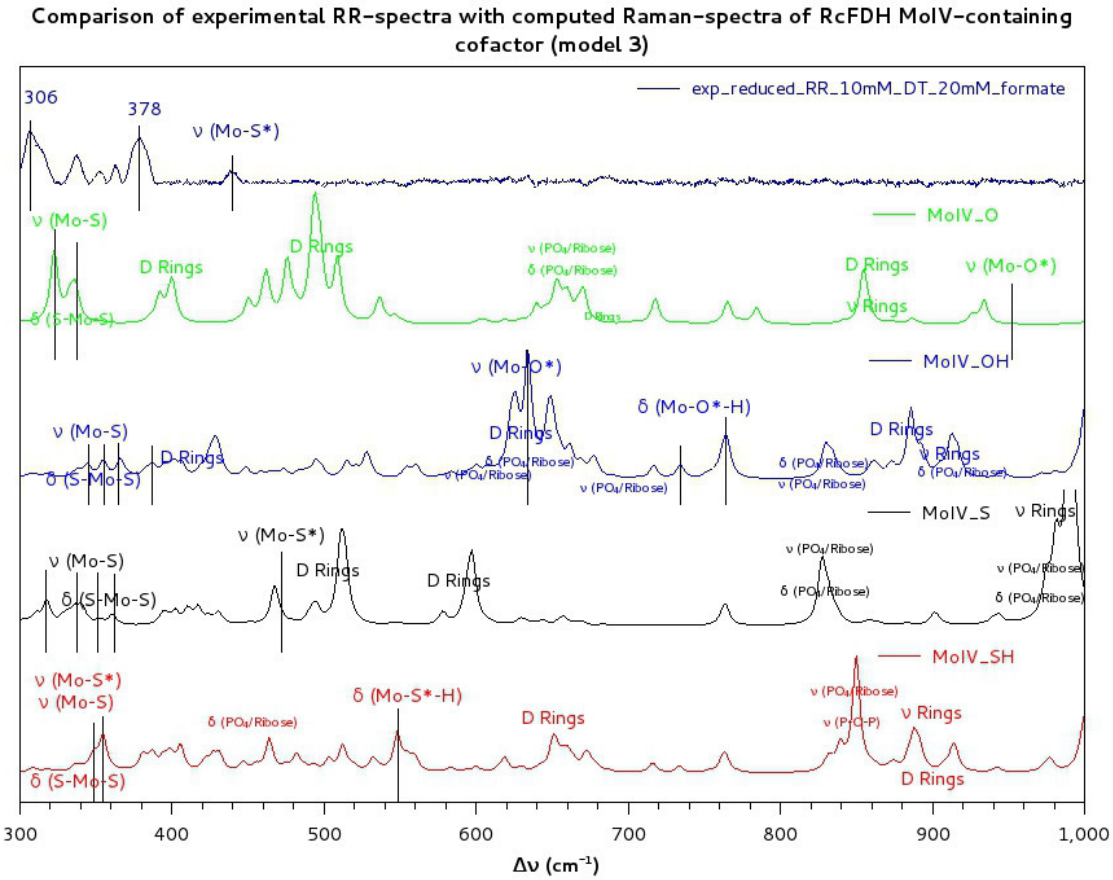

Figure A4. Calculated Raman-spectra of model 3 compounds (ε=1) for different ligand binding motifs at pentacoordinated MoIV (green: MoIV_O*, blue: MoIV_OH, black: MoIV_S*, red: MoIV_SH) without explicit treatment of Cys-ligand and comparison with experimental RR-spectrum: dark-blue: reduced form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73])

In both model 1 and 3 compounds with SH-ligation at the MoIV (with ε=1) signals near 440 cm^{-1} can be observed. Therefore, it cannot be completely ruled out that a MoIV_SH state might be present in the catalytic cycle. It may be that this species is an intermediate after formal hydride-transfer from formate which is only short-living. This has to be further investigated. Nevertheless, the presence of a protonated Cys in near vicinity to MoIV with ligated sulfido-group yielded the best agreement to the experimental signal as shown in section 13.3.1. Thus it is most likely that a pentacoordinated MoIV_S* model is the best representation of the real ligation-site at the Moco. Furthermore, H/D exchange-experiments [73] yielded no measurable shift of the band at 439 cm^{-1}.

B. Raman-spectra computations of model 1 compounds of Moco in *Rc*FDH in vacuum (ε=1) compared to Raman-spectra of model 1 compounds in implicit protein environment (ε=4)

In model 1 compounds the GDP-units of the native Moco of *Rc*FDH are completely omitted. For detailed model-composition please consult section 5.3.

Oxidized form of Moco (MoVI)

Comparison between computed Raman-spectra either with ε=1 or ε=4:

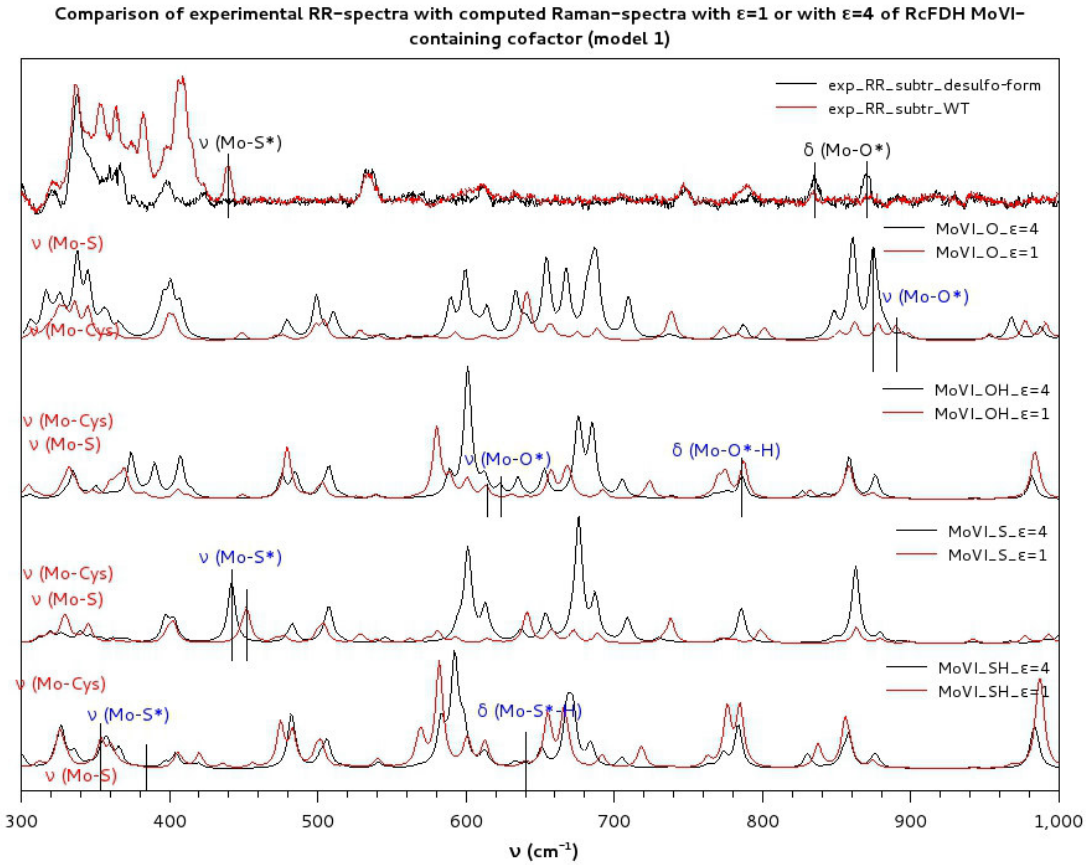

Figure B1. Calculated Raman-spectra of model 1 compounds for different ligand binding motifs at hexacoordinated MoVI (either MoVI_O*, MoVI_OH, MoVI_S* or MoVI_SH): black: ε=4, red: ε=1 and comparison with experimental subtractions RR-spectra of WT (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

Cys-ligation in MoVI of *Rc*FDH Moco (comparison to calculations with ε=4):

	ν (Mo-S) (cm^{-1})	ν (Mo-Cys) (cm^{-1})	δ (S-Mo-S) (cm^{-1})	δ (Mo-O-H) (cm^{-1})	δ (Mo-S-H) (cm^{-1})	ν (Mo-S*) (cm^{-1})	ν (Mo-O*) (cm^{-1})
Exp. Subtr. RR (WT)	?	?	?	-	?	439	-
Exp. Subtr RR (prestate)	?	?	?	?	-	-	830/871
MoVI_O* without Cys (model 1 + ε=4)	361-382	-	176	-	-	-	953
MoVI_O* without Cys (model 1 + ε=1)	379/386	-	179	-	-	-	973
MoVI_OH without Cys (model 1 + ε=4)	341/373	-	133/181	441/700	-	-	700
MoVI_OH without Cys (model 1 + ε=1)	342/383	-	171	411/731	-	-	411/731
MoVI_S* without Cys (model 1 + ε=4)	360-382	-	176	-	-	519	-
MoVI_S* without Cys (model 1 + ε=1)	354/376	-	175	-	-	538	-
MoVI_SH without Cys (model 1 + ε=4)	352-394	-	102	-	514	394	-
MoVI_SH without Cys (model 1 + ε=1)	373/412	-	126	-	524	412	-

Table B1. Vibrational frequencies for marker bands of pentacoordinated model 1 compounds, all without Cys, differing in ligand (O*, OH, S* or SH) and possible usage of implicit protein environment (either ε=1 or ε=4). compared to experimental subtraction RR-spectra (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

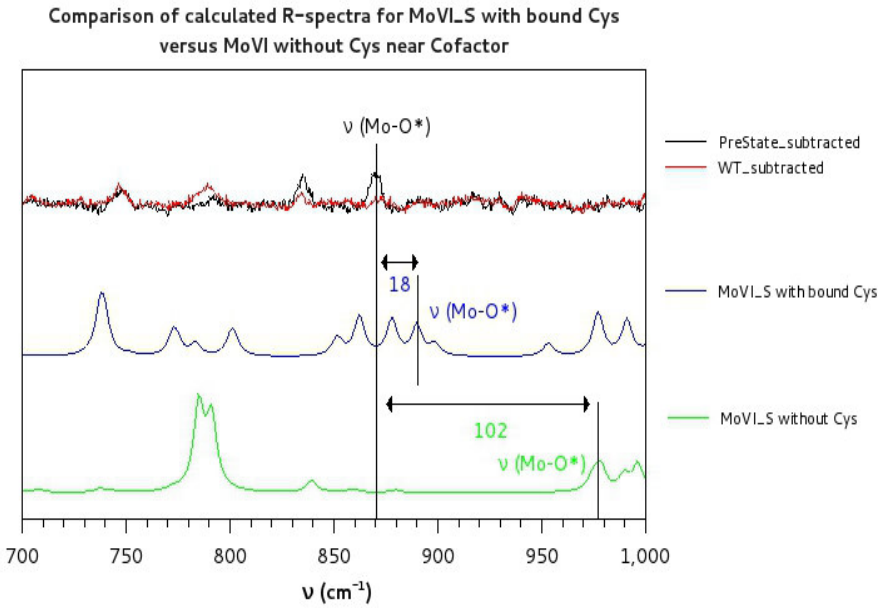

Figure B2. Calculated Raman-spectra for MoVI_O* model 1 compounds (ε=1): with (blue) or without (green) Cys and comparison to exp. subtraction RR-spectra: red: WT, black: desulfo-form (-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

Pentacoordinated MoVI model 1 compounds (both with ε=1 and ε=4) show a better agreement of the models with SH-ligation in terms of a Mo-S* stretching vibration to the experimental signal at 439 cm^{-1} than the models with S*-ligation. But, hexacoordinated MoVI_S* models yield an even better agreement to the experiment. Also other experimental studies (such as XAS and kinetic assays [49],[50],[54]) assume a sulfido-ligand at a hexacoordinated MoVI. Furthermore, SH-ligation at MoVI always led to a more pentacoordinated structure of the Moco, even in the presence of Cys (as discussed in more detail in Appendix G). But X-ray studies [52] of the FDH of *E. coli,* as well as computational investigations, such as done by Mota et al. [53] showed the oxidized form of the Moco to have a hexacoordinated trigonal-prismatic structure. Further, H/D-exchange experiments [73] showed no downshift. Therefore, it is very unlikely that a SH-ligation is present at MoVI.

H/D exchange (comparison between calculations with ε=1 and ε=4):

	δ (Mo-O-H)(cm^{-1})	**Δν (cm^{-1})**
Exp. Subtr RR (prestate)	830/871	
Exp. Subtr RR (prestate) → H/D exchange	830/873	+ 2
MoIV_OH model 1 (bound Cys) + ε=1	770/787	
MoIV_OD model 1 (bound Cys) + ε=1	558/638	- 212/-149
MoIV_OH model 1 (bound Cys) + ε=4	767/786	
MoIV_OD model 1 (bound Cys) + ε=4	566/627	-201/-159

Table B2. Calculated vibrational frequencies for hexacoordinated MoVI_OH and MoVI_OD model 1 compounds (either ε=1 or ε=4) and comparison with experimental RR-spectra (H/D exchange; 1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

Disulfide-bridge formation (comparison to calculations with ε=4):

	ν (Mo-S*) (cm^{-1})	ν (S*-S) (cm^{-1})	δ (Mo-S*-S) (cm^{-1})	δ (S*-S-C) (cm^{-1})
Exp. Subtr. RR (WT)	439	?	?	?
Exp. Red. RR	439/438	?	?	?
MoVI_S*-S (model 1 + ε=1)	406	470/572/721	406/470/572	470/572/721
MoVI_S*-S (model 1 + ε=4)	398	489	489/595	489/595/712
MoIV_S*-S (model 1 + ε=1)	385	416/585	416/585/757	416/757/945
MoIV_S*-S (model 1 + ε=4)	388	423/585	423/585	423/757/946

Table B3. Calculated vibrational frequencies for pentacoordinated MoVI and MoIV model 1 compounds (with either ε=1 or ε=4) with explicit modeling of disulfide bond between sulfido and Cys-ligand (MoVI/IV_S*-S) and comparison to corresponding experimental RR-spectra (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 10 mM formate; provided by Stefan Wahlefeld [73])

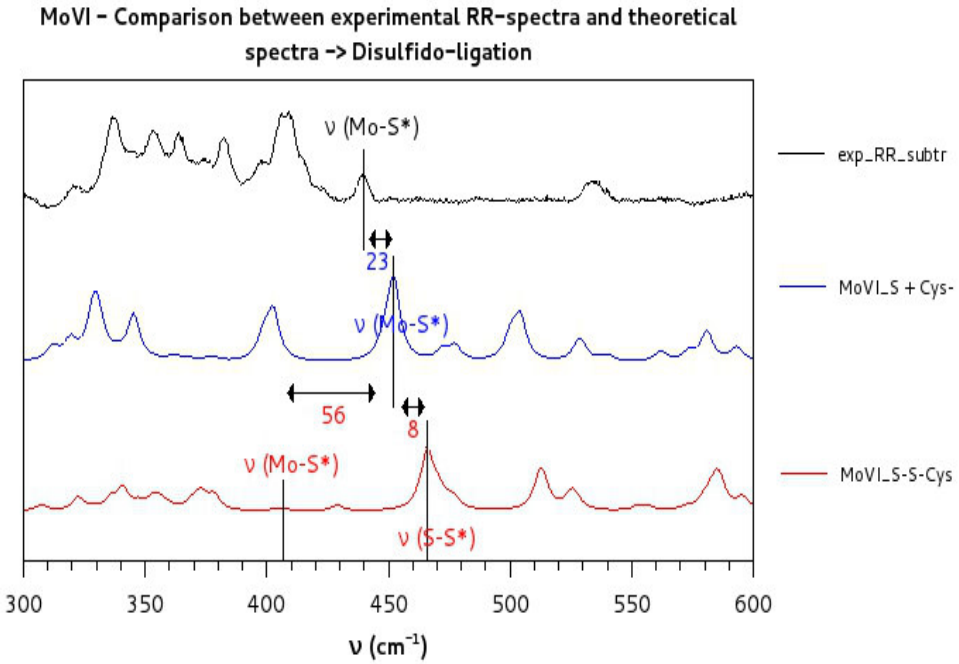

Figure B3. Calculated Raman-spectra for MoVI model 1 compounds (ε=1): with bound S*- and Cys-ligand (hexacoordinated Mo; MoVI_S*; blue); with explicit treatment of disulfide bond between sulfido and Cys-ligand (pentacoordinated Mo; MoVI_S*-S; red) and comparison to experimental subtraction RR-spectrum of WT (black) (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; provided by Stefan Wahlefeld [73])

Reduced form of Moco (MoIV)

Comparison between computed Raman-spectra of model 1 compounds either with ε=1 or ε=4:

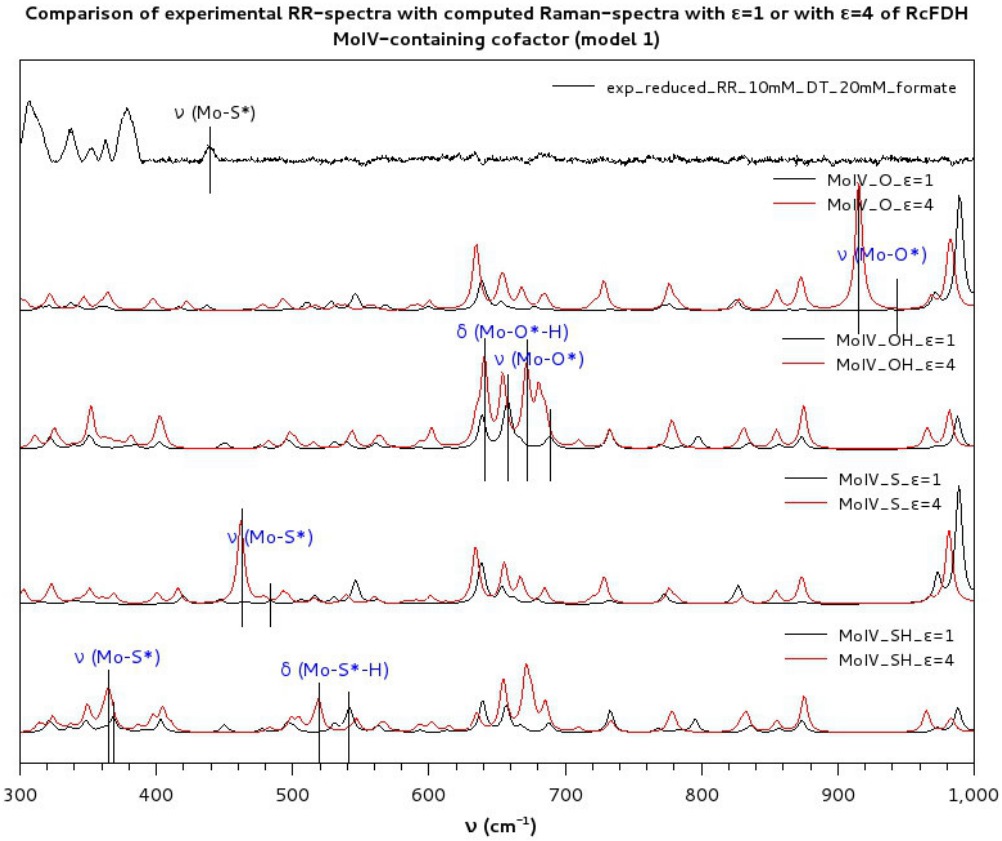

Figure B4. Calculated Raman-spectra of model 1 compounds for different ligand binding motifs at pentacoordinated MoIV (MoIV_O*, MoIV_OH, MoIV_S* or MoIV_SH): red: ε=4; black: ε=1; and comparison with experimental RR-spectrum: black: reduced form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 20 mM formate in 10 mM DT; provided by Stefan Wahlefeld [73])

Different possible ligation-sites for Cys (comparison to calculations with ε=4):

	ν (Mo-S) (cm^{-1})	δ (S-Mo-S) (cm^{-1})	ν (Mo-S*) (cm^{-1})	ν (S*-S) (cm^{-1})	ν (Mo-Cys) (cm^{-1})
Exp. Red. RR	?	?	439	?	?
MoIV_S* (model 1 + ε=1)	319/352	352/483	176/483	-	-
MoIV_S* (model 1 + ε=4)	322-360	111/182	461	-	-
MoIV_S*_Cys^- (model 1 + ε=1)	-	-	188/425	425	-
MoIV_S*_Cys^- (model 1 + ε=4)	-	-	359/369	359/369	339/359/369
MoIV_S*_CysH (model 1 + ε=1)	-	-	456/468	-	-
MoIV_S*_CysH (model 1 + ε=4)	-	-	448	-	-
MoIV_S*_bound_Cys^- (model 1 + ε=4)	-	-	405	405	186/229-279

Table B4. Calculated vibrational frequencies for MoIV_S* model 1 compounds with different ligation sites of Cys-ligand (pentacoordinated Mo; Cys unbound to MoIV, but either charged (MoIV_S*_Cys^-) or protonated (MoIV_S*_CysH), or hexacoordinated Mo with Cys bound to MoIV (MoIV_S*_bound_Cys^-)) with either ε=1 or ε=4 and comparison to corresponding experimental RR-spectrum (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 10 mM formate; provided by Stefan Wahlefeld [73])

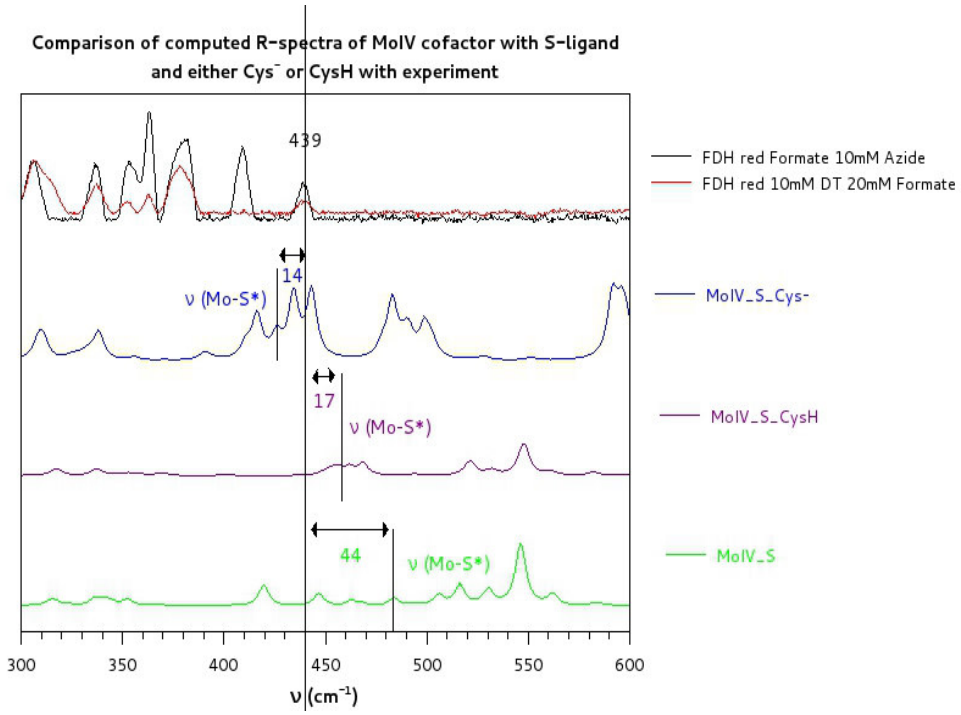

Figure B5. Calculated Raman-spectra in vacuum for MoIV_S* model 1 compounds (ε=1) with different ligation sites of Cys-ligand (pentacoordinated Mo, Cys not ligated to MoIV, but either charged (MoIV_S*_Cys^-; blue) or protonated (MoIV_S*_CysH; violet) or completely absent (MoIV_S*; green)) and comparison to corresponding exp. RR-spectra of reduced form (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with either 10 mM formate (black) or 20 mM formate in 10 mM DT (red); provided by Stefan Wahlefeld [73])

Disulfide-bridge formation (ε=1):

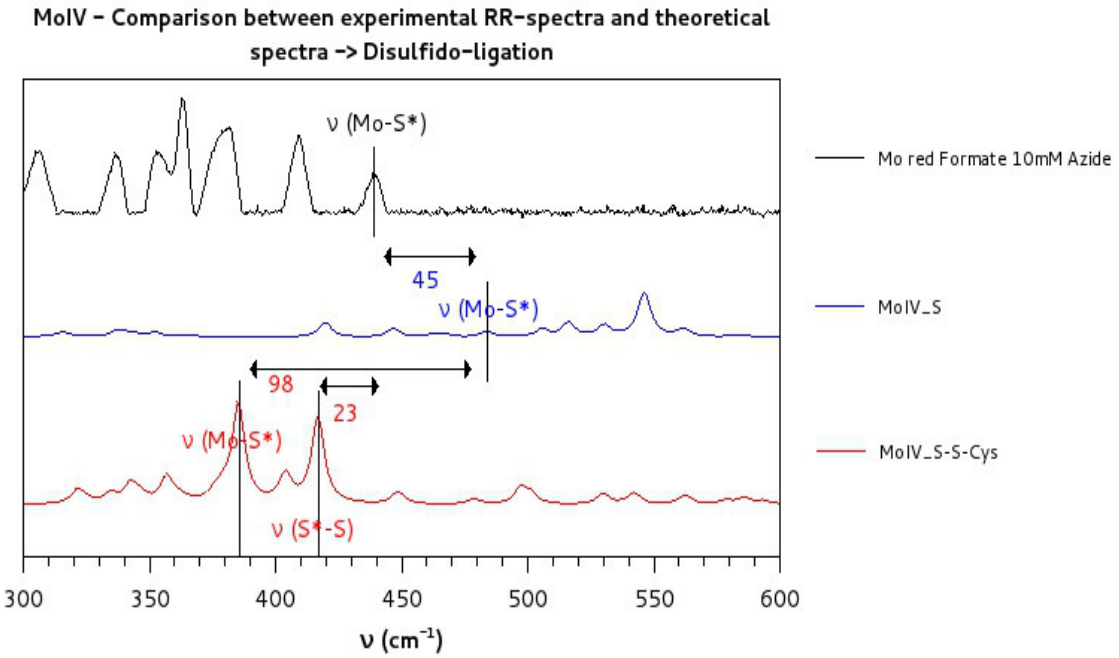

Figure B6. Calculated Raman-spectra for pentacoordinated MoIV model 1 compounds (ε=1) with S*- and no Cys-ligand (MoIV_S*; blue) or explicit treatment of disulfide bond between sulfido and Cys-ligand (MoIV_S*-S; red) and comparison to experimental RR-spectrum of reduced form (black; 1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide; reduction with 10 mM formate; provided by Stefan Wahlefeld [73])

C. Different possible azide binding-sites at MoVI of *Rc*FDH Moco for model 1 compounds (ε=1 , ε=4 and ε=80):

Purification processes of FDH are often carried out in the presence of a stabilizing salt which in most cases is azide (N_3^-). Azide is also known to be an inhibitor to the formate oxidation (e.g. [55]). If the mode of inhibition and how azide interacts with molybdenum or the Moco in general, would be known, it might help to further understand the possible Moco-involvement in formate-oxidation.

Azide binding-modes

The possible binding sites of azide are investigated in three different models, schematically shown in Figure C1. All models used the model 1 structure (omitting of GDP's), with a sulfido-ligand present and (a) a directly bound azide to MoVI (MoVI_S*_N_3^-; charge of -1 assumed for azide) and no Cys present, (b) Cys still ligated to MoVI and azide forming a covalent bond with the S*-ligand (MoVI_S*-N_3; charge of -1 assumed for the S-N_3-ligand) and (c) sulfido and Cys directly coordinated to the metal with an unbound negatively charged azide ion 5 Å away from metal (MoVI_S*_unbound_N_3^-; charge of -1 for azide ion). For the charges of the other parts of the model 1 compounds and model composition see section 5.3. The three models have been optimized geometrically (with PCM solvation methods and ε=1, ε=4 or ε=80) and subsequent vibrational frequencies have been computed and compared to experimental ones (IR- and RR-experiments) [73]. The results are shown in Table C1.

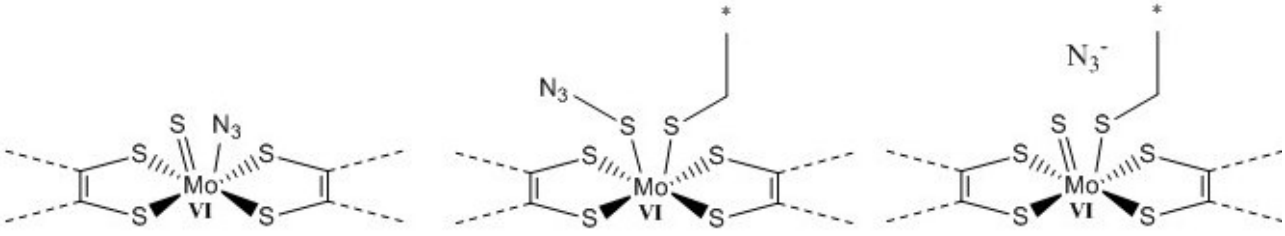

Figure C1. Schematic visualization of different N_3^- binding sites (dotted lines represent pterin units which are not shown here) for model 1 compounds: (a) MoVI_S*_N_3^- (model-charge: -1); (b) MoVI_S*-N_3 (model-charge: 0; only computed with ε=4); (c) MoVI_S*_undbound_N_3^- (model-charge: -2) (from left to right)

The results for the N=N vibrations of azide for the models in vacuum show strong deviations from the experimental frequency at 2031 cm^{-1}. The model (a) with directly coordinated azide to the metal shows the biggest shift of about $\Delta\nu$=134 cm^{-1}, while the one with unbound ion shows the least deviations from the experimental band of about $\Delta\nu$=46 cm^{-1}. This model also shows a temporary formation of a bond between azide and the S*-ligand during the optimization procedure. Further no direct coordination to the metal has been observed during the optimization of model (c). The Mo-S* stretching is shifted to lower values in the case of an unbound ion in (c) of about $\Delta\nu$=44 cm^{-1} and is in less good agreement to the experiment than in the model with bound azide (a). But this may be due to the missing simulation of protein environment. The findings concerning the N=N-vibrations indicate that the azide is not binding directly to MoVI. The ion stays near the metal site of the cofactor and thus may be blocking the entryway for formate by steric or electronic influence. The formation of a S*-N_3 bond is also possible and may hinder the oxidation of $HCOO^-$, since S* plays an important role as hydride-acceptor in the catalytic cycle, as pointed out in the work of Maia et al. [54] and as discussed in sections 13.6. and 13.7. To get an impression of the influence of the environment on the computed frequencies, the three named azide binding models have also been computed with usage of implicit protein environment (ε=4). In Table C1 the obtained vibrational frequencies are depicted.

	ν (Mo-S*) (cm^{-1})	shift from exp. Δν (cm^{-1})	ν (N_3)(cm^{-1})	shift from exp. Δν (cm^{-1})
Exp. subtraction RR-spectra (WT)	439		2031	
ε=1				
MoVI_S*_N_3^- (model 1 + ε=1)	473	+ **34**	2165	+ **134**
MoVI_S*_unbound_N_3^- (model 1 + ε=1)	395	**-44**	2077	+ **46**
ε=4				
MoVI_S*_N_3^- (model 1 + ε=4)	458	**+19**	2128	**+97**
MoVI_S*-N_3 (model 1 + ε=4)	417	**-22**	2014	**-17**
MoVI_S*_unbound_N_3^- (model 1 + ε=4)	369	**-70**	2043	**+12**

Table C1. Calculated vibrational frequencies for different N_3^- binding-modes at the MoVI site; model 1 compounds according to Figure C1 (with either ε=1 or ε=4)) and comparison to experimental RR-data (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide) and IR-data; provided by Stefan Wahlefeld [73]

The overall trend in computations with ε=4 is the same as for ε=1: The lowest shift for N=N vibrations from the experimental data [73] appears in model (c) MoVI_S*_unbound_N_3^-, with unbound azide, at 2043 cm^{-1} (deviation of Δν=12 cm^{-1} from experiment). Unfortunately the Mo-S* vibration at 369 cm^{-1} is strongly displaced from the observed frequency in the measurements about Δν=70 cm^{-1}. In contrast to this, the model (a) MoVI_S*_N_3^- with metal-bound azide yields the highest displacement from the measurements for the azide vibration. Here, the corresponding azide-stretching mode could be observed at 2128 cm^{-1} (difference to exp.: Δν=97 cm^{1}). In model MoVI_S*-N_3, in which the azide is linked to the metal by formation of a S-N_3-bond, the azide stretching vibration occurs at 2014 cm^{-1} (shifted from exp. about Δν=17 cm^{-1}) and is in good agreement to the IR-data (similar as in MoVI_S*_unbound_N_3^-). Furthermore, the Mo-S* stretching in MoVI_S*-N_3 at 417 cm^{-1} (difference to exp.: Δν=22 cm^{-1}) is also near to the experimental one at 439 cm^{-1}. Additionally, the formation of a S-N_3 bond can be seen in the calculations with unbound ion in model (c). This and the very good agreement of the Mo-N_3 stretching in models (b) and (c) leads to the same conclusion as for the computations in vacuum: The azide ion does not bind directly to the metal center and stays either unbound near MoVI or forms a bond with the sulfido-ligand, thus hindering formate approaching the active site. This is in perfect agreement to the results of bond-breaking energy computations for hexacoordinated MoVI with bound sulfido-group (s. section 13.5.). These yielded a very stable MoVI-Cys bond.

Comparison with experimental findings

These findings indicate that azide may inhibit the active site of *Rc*FDH *via* direct interaction with the S*-ligand or blocking the dissociation of Cys in MoIV (s. sections 13.5. to 13.7.) Furthermore, also the inhibition of the enzyme *via* blocking the entryway for the substrate into the active-site is possible. But since the cofactor-binding pocket has not been simulated in greater detail here, this hypothesis could not be investigated more closely.

These results are in direct contrast to recent findings of Robinson et al. [55]: They investigated the effect of different inhibitor molecules (like azide) on the catalytic activity of both formate oxidation and carbon dioxide reduction of formate dehydrogenase H from *Escherichia coli* (*Ec*FDH-H). They found that N_3^- inhibits formate oxidation and that azide is competitive to the substrate formate. It only weakly inhibits CO_2-reduction. They discussed that azide selectively interacts with the MoVI-state. For other potential inhibitor species they found a similar selectivity to an oxidized molybdenum. Therefore, they proposed a direct coordination of azide to the molybdenum (upon release of Cys) or an interaction of azide with the sulfido-ligand. This they found further established by investigations of the net ligand electron donation (ligand electrochemical parameter E_L [201]) and the ligand π-electron donating ability (ligand resonance field donation [202]) of different inhibitors in comparison to inhibitor-dissociation constants: Strong net and π-electron donors are forming stable inhibitor-Moco complexes and showing low dissociation-constants. Since MoVI is electron deficient, they concluded that the inhibitor species (like azide) are directly ligating to the molybdenum (forcing Cys to leave it coordination position) and further donating π-electron-density into the d-orbitals of Mo. Therefore, they favored this mode of action over inhibitor-sulfido-interaction.

Because of the contrasting results by Robinson et al. [55] it has to be verified that the differences between computations and experiments shown here are not due to the choice of the wrong theoretical method. Therefore additional calculations have been carried out: First, frequencies for the isolated azide ion solvated in water by application of a dielectric field with ε=80 have been computed. Second, the MoVI_S*_N_3^- model (a) with bound azide has been recomputed with a dielectric of ε=80. The results can be seen in Table C2:

The obtained N=N vibrational frequency for the free azide at 2018 cm^{-1} is in good agreement to the IR-data (exp.: 2049 cm^{-1}), with only a difference of $\Delta\nu$=31 cm^{-1}. The computations of the model with bound azide to MoVI (ε=80) yields a frequency at 2111 cm^{-1} (difference to exp.: $\Delta\nu$=67 cm^{-1}) which is nearly half as far away, as for computations with ε=1 (s. Table C1 and C2). Both the azide stretching, as well as the Mo-S* stretching, are in better agreement to the experimental frequencies than even the results for calculations with ε=4. Because of the differing accuracy of the computed frequencies, when changing the environment of the molecule, it is evident that the vibrational modes of azide are very sensitive to environmental/solvation effects. Thus, the differences between computed and experimental vibrational frequencies for the N=N streching modes are due to the treatment of the environment of the Moco-models. Since changing the environment/solvation leads to increased accuracy of the computed vibrations, the applied method (DFT, BP86 functional) is sufficient for the calculations. It is not causing the differences in vibrational frequencies, since otherwise, the environment should not have such an influence on the vibrations. Thus the here presented results are reliable.

As discussed here, it is very probable that azide is not binding directly to MoVI. Additionally, investigations of metal-independent (e.g. yeast) formate dehydrogenase show that substrate oxidation is inhibited by azide, although no molybdenum is present [203],[204]. This may be due to interaction of the inhibitor with conserved protein residues Arg and His [55],[203] which can also be found in *Rc*FDH and *Ec*FDH-H in near vicinity of the Moco. Computations of models with inclusion of these residues would shed further light on the matter. It is also possible that the binding of Cys to the MoVI-center is energetically more stable than a possible MoVI-N_3 binding, as it is in comparison to formate-metal ligation (s. section 13.5.). Evaluation of bond-breaking enthalpies for a MoVI-N_3 bond might be necessary to investigate this. Furthermore, an interaction of azide with the sulfido-ligand at MoVI can occur *via* π-electron-donation to the empty d-orbitals of the S*-ligand. This would also be in agreement with the experimental findings of Robinson et al. [55]

concerning the connection between inhibitor dissociation constant and electron donation ability of the inhibitor. The selectivity of the inhibition of the MoVI-state might be explained by different charge-densities in the active center in different oxidation states of Mo. Since MoIV carries more electronic charge, also S* might carry more charge and the d-orbitals are not as available for azide-sulfido interaction, as in MoVI. Also possible protein rearrangements, could make azide at the MoIV-site more unstable. X-ray experiments would provide more insight.

	ν (Mo-S*) (cm^{-1})	**shift from exp. Δν (cm^{-1})**	**ν (N_3^-)(cm^{-1})**	**shift from exp. Δν (cm^{-1})**	**ν (free N_3^-)**	**shift from exp. Δν (cm^{-1})**
Exp. RR-spectra (WT)	439		2031		2049	
ε=80						
free N_3^- (ε=80)	-	-	-	-	2018	-31
MoVI_S*_N_3^- (model 1 + ε=80)	451	+12	2111	+80	-	-
ε=1						
MoVI_S*_N_3^- (model 1 + ε=1)	473	+34	2165	+134	-	-
ε=4						
MoVI_S*_N_3^- (model 1 + ε=4)	458	+19	2128	+97	-	-

Table C2. Calculated vibrational frequencies of either free N_3^- or model 1 compounds MoVI_S*_N_3^- with metal-bound N_3^- (ε=1, 4 or 80) and comparison to experimental RR-data (1-1.5 mM protein with 40% Moco loading in 100 mM Tris buffer with 10 mM azide) and IR-data; provided by Stefan Wahlefeld [73]

Possibly, the calculations could be redone with additional inhibitor-species (such as OCN^-, SCN^- and NO_3^-) and compared to each other. Also, potential protein interaction partners (such as His387 and Arg587) have to be included into the existing computational models. QM/MM computations of *Rc*FDH with different inhibitor species, followed by frequency-calculations of the active center with explicit treatment of water would probably yield the most accurate results.

According to the here presented results in combination with vibrational spectroscopic results from Wahlefeld et al. [73] azide is located in near vicinity of the Moco and may interact with the S*-ligand at the MoVI-site, but does not ligate directly to MoVI. This has also been suggested by EPR-measurements [54],[196]. Presence of N_3^- might render it difficult for formate to interact with the Moco, as well as making it impossible for Cys to leave its coordination-site due to sterical or energetic hindrance by the present azide. That would be why formate might not be able to undergo oxidation in the presence of azide. It is also possible that azide blocks the entryway of formate. In all cases formate-oxidation would be abolished. The here presented results are in agreement to the findings of an indirect interaction of formate with the Moco, as discussed in sections 13.5. to 13.7.

D. Energy calculations concerning bond-stability of Mo-Cys and Mo-OCHO bonds in Moco of *Rc*FDH (ε=1):

Similar computations as described in 13.5. have been carried out with models as depicted in Figure D1 in vacuum with ε=1.

Figure D1. Schematic representation of calculated MoVI and MoIV model 1 compounds with S*-ligand present and: (left) MoVI/IV_S*_bound_Cys^- (bound Cys); (right) MoVI/IV_S*_bound_$HCOO^-$ (bound formate)

The results are given in Table D1.

ε=1	**$\Delta H_{bond\text{-}breaking}$ (kcal/mol)**
MoVI_S*_bound_Cys^- (model 1 + ε=1)	+50.890
MoVI_S*_bound_$HCOO^-$ (model 1 + ε=1)	+38.291
MoIV_S*_bound_Cys^- (model 1 + ε=1)	-83.267
MoIV_S*_bound_$HCOO^-$ (model 1 + ε=1)	-39.084

Table D1. Calculated bond-breaking enthalpies for either Cys- or formate-ligation at both MoVI and MoIV sites (ε=1)

In comparison to 13.5. the bond-breaking enthalpies are all higher. An increased stability of the MoVI-HCOO bond could be observed which is nonetheless lower than the one for the MoVI-Cys bond. Dissociation of Cys from the metal in its oxidized form would require ~12 kcal/mol more energy, as the formation of the formate-metal bond would yield. In the reduced state both species are still not stably bound to MoIV, with dissociation of Cys yielding the highest energy. This trend is the same, as for the calculations discussed in 13.5.

E. Energy calculations concerning bond-stability of Mo-Cys and Mo-OCHO bonds (ε=1) with alternative models of Moco of *Rc*FDH:

Bond-breaking enthalpies have been computed for models with adjacent unbound ligand, as depicted in Figure E1. One molecule (either Cys^- or $HCOO^-$) is bound to the metal center, while the other has been placed unbound in 5 Å distance to the molybdenum. Enthalpies have than been computed according to the scheme shown in 13.5. and 5.3.2. Furthermore, all computations have been carried out in vacuum. The results are given in Table E1.

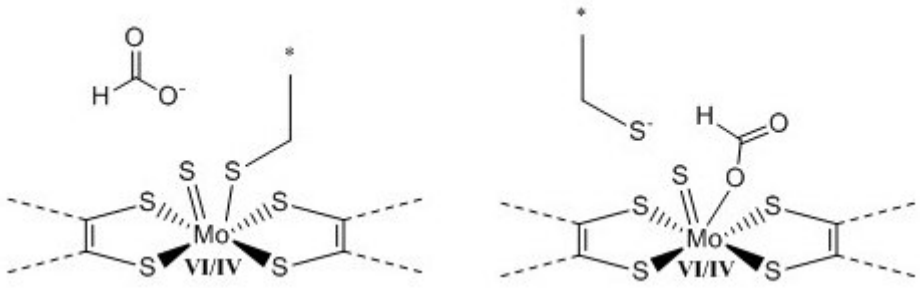

Figure E1. Schematic representation of calculated MoVI and MoIV model 1 compounds with S*-ligand present and: (left) MoVI/IV_S*_bound_Cys^-_unbound_$HCOO^-$ (bound Cys and unbound formate); (right) MoVI/IV_S*_bound_$HCOO^-$_unbound_Cys^- (bound formate and unbound Cys^-)

ε=1	**$\Delta H_{bond\text{-}breaking}$ (kcal/mol)**
MoVI_S*_bound_Cys^-_unbound_$HCOO^-$ (model 1 + ε=1)	+1.099
MoVI_S*_bound_$HCOO^-$_unbound_Cys^- (model 1 + ε=1)	-17.745
MoIV_S*_bound_Cys^-_unbound_$HCOO^-$ (model 1 + ε=1)	-97.211
MoIV_S*_bound_$HCOO^-$_unbound_Cys^- (model 1 + ε=1)	-7.406

Table E1. Calculated bond-breaking enthalpies for either Cys- or formate-ligation at both hexacoordinated MoVI- and MoIV-sites (ε=1)

The enthalpies are all lower than the ones without adjacent ion (as discussed in 13.5. and Appendix D). This has to be due to the influence of the unbound ligand. In contrast to the results presented in section 13.5. and Appendix D, it is shown in Table E1 that the dissociation of formate from MoVI would even yield energy (-17.745 kcal/mol). This further underlines the assumption that formate is not directly ligating to the metal center of the Moco in its oxidized form and thus has to interact with it *via* the sulfido group. Similar to the bond-breaking enthalpies as shown in 13.5., Cys is bonded more strongly to MoVI (1.099 kcal/mol are needed to break the Mo-Cys bond) than formate (difference between bond-breaking enthalpies for MoVI-Cys and MoVI-OCHO is about $|\Delta(\Delta H_{bond\text{-}breaking})|$=18.844 kcal/mol). In the reduced MoIV-form it is *vice versa*: Both ligands do not bind to MoIV: Formate yields an energy of 7.406 kcal/mol upon dissociation from MoIV, while Cys yields 97.211 kcal/mol. This is even more energy than observed for all other enthalpy-computations as discussed in this work and confirms the proposal that Cys is leaving MoIV. Formate is probably not binding MoIV, since energy is needed for this. Thus, a pentacoordinated MoIV is created. The observed trend is maintained: Cys only leaves the metal, if molybdenum is in its +IV form.

F. Alternative mechanism of catalytic oxidation of formate *via* direct formate-metal ligation at Moco of *Rc*FDH

As pointed out in section 13.5. (and also depicted in Tables D1 and E1), the difference in bond-breaking enthalpies for both MoVI-Cys and MoVI-HCOO ligations is small enough to be possibly compensated by changes in the protein environment. This could render a direct formate-binding to MoVI possible. For the sake of completeness, a possible mechanism including direct ligation of formate to the MoVI is depicted in Figure F1. This is based on the comparisons between computed Raman-spectra with experimental RR-data [73], as discussed in part V of this work. Arg587 and His387 contributions are according to results from Hartmann et al. [49], while EPR-results from Dufus et al. [73],[196] attributed to MoV-state contributions.

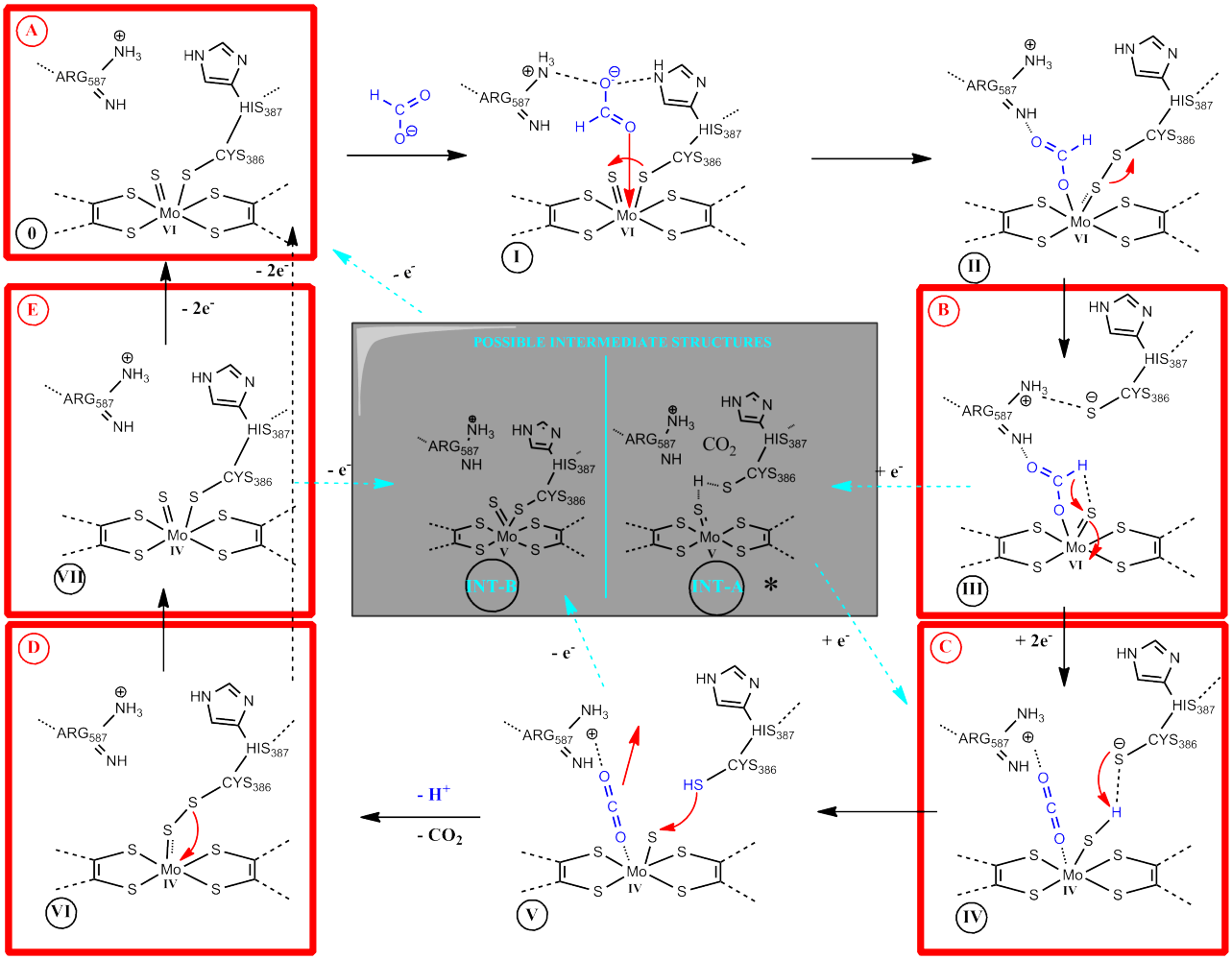

Figure F1. Proposed catalytic cycle for reversible formate oxidation at the Moco-site using available experimental [49],[73],[196] and theoretical data and including hydride-/proton-transfer from formate to sulfido-ligand, disclocation of Cys^- from the metal, direct liagtion of formate to MoVI and sulfur-shift mechanistic parts, as well as possible MoV-intermediates (* INT-A with inclusion of CO_2 and its possible interaction with the metal has not been tested)

It starts with a hexacoordinated MoVI-species (**0**). Approaching formate is directly ligating to the metal, forcing Cys to dissociate from MoVI (**I**) (maybe *via* disulfide-bridge formation; **II**). The hydrogen of the substrate is interacting with the S*-ligand (**III**), ultimately resulting in a formal hydride- or proton-transfer and the formation of a SH-ligand (**IV**). Further two electrons are translocated to the metal yielding the reduction of MoVI to MoIV (**IV**). Deprotonated Cys abstracts a proton from the SH-group (**V**). The proton is further transferred in the protein matrix and Cys

returns to the metal *via* disulfide-bridge formation (**V** + **VI**) while the metal is still in its reduced state MoIV (**VII**) or Cys returns after MoIV is back oxidized to MoVI (from **VI** directly to **0**). The available data on this particular point is not completely clear: RR-data, as well as computed Raman-spectra suggest the possibility of a hexacoordinated MoIV-species with bound Cys (see section 13.3.1. and 13.6.), while bond-breaking enthalpy calculations show the energetically unfavorable MoIV-Cys ligation (see section 13.5. and 13.6, as well as Appendix D + E). CO_2 is leaving the active-site. Two electrons are translocated into the protein (MoIV → MoIV). The cycle can start again with a hexacoordinated MoVI-species (**0**).

Also a stepwise electron-transfer from formate to the molybdenum *via* MoV (MoVI → MoV → MoIV and *vice versa*) is possible by three different routes: The electrons can be transferred from **II** *via* MoV-species (**INT-A**) to result in **IV** and it is possible that the back-oxidation of the cofactor is happening by simultaneous back-binding of Cys and step-by-step electron-translocation (from **V** via **INT-B** with MoV to **0**). It is also possible that Cys is primary returning to MoIV and afterward two electrons are transferred stepwise (from **VII** via **INT-B** with hexacoordinated MoV to **0**).

In general, due to the stable MoVI-Cys bond, it is unlikely that Cys is dissociating from MoVI and is exchanged by formate. In contrast to this, only ~6 kcal/mol have to be compensated for direct ligation of formate to MoIV (see 13.5. and Appendix E). And since in the MoIV state Cys is dislocating from the metal a free coordination position for formate would be available. A possible mechanism, including the direct ligation of $HCOO^-$ to MoIV is depicted in Figure F2. Again contributions of Arg587 and His387 are due to results from Hartmann et al. [49].

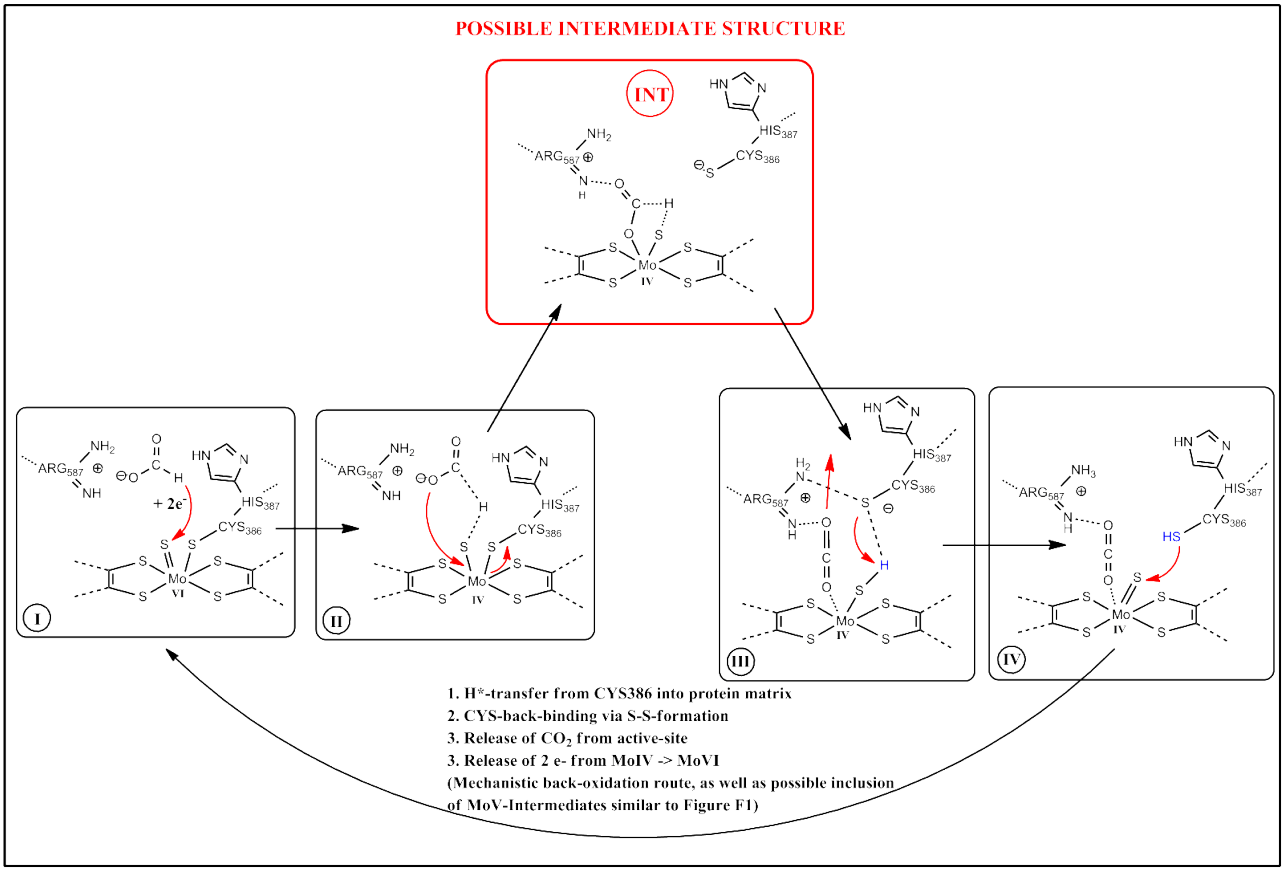

Figure F2. Proposed catalytic cycle for reversible formate oxidation at the Moco-site using available experimental [49],[73],[196] and theoretical data and including hydride-/proton-transfer from formate to sulfido-ligand, dislocation of Cys^- from the metal, direct ligation of formate to MoIV state; the steps back from IV to I, as well as possible MoV-states are not included and are similar to steps VI to 0 and INT, respectively, in Figure F1.

Similar to Figure F1, the cycle starts with a hexacoordinated MoVI state (all-sulfur coordination). Upon approach of formate (I) two electrons are transferred (maybe via formal hydride) to MoVI which is then reduced to MoIV (II). The hydrogen of formate interacts with the sulfido-ligand, forming a thiol-group. After reduction of the molybdenum, Cys is leaving the coordination-site, thus creating a vacant position at the metal. Now, formate can directly bind to the metal-center (the necessary energy has to be generated by protein rearrangements or energy-gains in other steps of the cycle) and thus an intermediate state is formed, in which a hexacoordinate MoIV with a thiol-ligand and a coordinated formate/CO_2 is created (INT). After complete transfer of the hydrogen from formate to the S*-ligand (III), carbon dioxide is leaving the active site. Deprotonated Cys is abstracting a proton from the newly formed thiol-group. Thus a pentacoordinated MoIV state is generated with a sulfido-ligand being present (IV). From here on, the reaction cycle follows the same procedure, as discussed in section 13.6 and displayed in Figure 75 and Figure F1: Cys returns *via* disulfide bridge-formation to the MoIV or MoVI states, after releasing a proton further into the protein matrix. In addition two electrons are also transferred from MoIV and the original hexacoordinated MoVI state (I) is regained. Participation of MoV-states is similar to Figures 75 and F1.

The existence of a hexacoordinated MoIV state with directly ligated formate has yet to be proven. It would be in agreement with findings of Mota et al. [53] and Schrapers and Hartmann et al. [49], [50]. But it contradicts the very good agreement of computed Raman-spectra for a pentacoordinated MoIV-species with experimental RR-data (see section 13.3.1.). This could mean that such an intermediate species is only shortliving. Further studies such as Raman-spectra computations for models according to INT in Figure F2 and comparison to RR-experiments have to be carried out to either verify or exclude this option. The inclusion of further protein residues into the existing models and subsequent repetition of bond-breaking enthalpy- and Raman-spectra-calculations might also shed further light onto this matter.

G. Analysis of structures of MoVI model 1 compounds of Moco in *Rc*FDH after geometry optimization:

In part V of this work the influence of different ligations at Mo (VI, V, IV) on the vibrational frequencies has been the focus. Here the geometries of four different hexacoordinated MoVI model 1 compounds (either O*-, OH-, S*- or SH-ligation, with Cys^- additionally ligated to MoVI; named accordingly MoVI_O*, MoVI_OH, MoVI_S* and MoVI_SH) after structural optimization with QM-methods (as discussed in section 5.3.) are investigated further.

Interesting characteristics could be found regarding the ligation of different ligands influencing the geometry of the cofactor. Bond-lengths and -angles of the first coordination sphere of MoVI have been investigated and compared between individual models and with the used homology model [49]. The results are presented in Table G1 and Figure G1. The numbers in parenthesis are according to the atom-numbering of Gaussview, as depicted in Figure G1.

X=(O, S)	MoVI_O*	MoVI_OH	MoVI_S*	MoVI_SH	Homology Model(MoVI_S*)
Bonds (Å)					
Mo(59)-S(4)Cys	**2.4348**	**2.641**	**2.413**	**2.584**	2.30858
Mo(59)-S(7)	2.4246	2.374	2.436	2.357	2.49953
Mo(59)-S(8)	2.5123	2.466	2.446	2.439	2.56997
Mo(59)-S(33)	2.4628	2.337	2.447	2.341	2.38523
Mo(59)-S(34)	2.4636	2.348	2.458	2.347	2.62953
Mo(59)-X(60)	**1.7333**	**1.925**	**2.202**	**2.375**	2.01123
Angles (°)					
S(4)Cys-Mo(59)-X(60)	**91.8**	**77.34**	**93.14**	**75.26**	75.71716
S(7)-Mo(59)-X(60)	**81.62**	**117.86**	**80.796**	**114.98**	98.68113
S(34)-Mo(59)-X(60)	91.35	91.38	88.54	89.95	104.762
S(4)Cys-Mo(59)-S(7)	**105.2**	**95.8**	**104.67**	**95.5**	74.60329
S(7)-Mo(59)-S(8)	86.495	80.88	79.55	81.76	74.01451
S(33)-Mo(59)-S(34)	79.84	83.84	79.59	82.51	81.0199
S(7)-Mo(59)-S(34)	**96.87**	**81.65**	**97.49**	**82.9**	68.7619
S(8)-Mo(59)-S(33)	**76.74**	**80.62**	**76.31**	**80.98**	94.60366

Table G1. Structural parameters (bond-lengths and -angles) of first coordination sphere of hexacoordinated MoVI model 1 compounds with O*-, OH-, S*- or SH-ligation after QM-geometry optimizations with ε=4 (numbering according to QM-model, as depicted in Figure E1), compared with data from homology model [49] with hexacoordinated MoVI with sulfido- and Cys-ligation; X denotes either O*-, OH-, S*- or SH-ligand, while S(4) is the sulfur of the modeled Cys

First, the presence of an O*-ligand at MoVI in contrast to S*-ligation produces an increase in the studied bond-lengths of the dithiolene- and Cys-ligations of the metal. The Mo-O* bond is significantly shorter (1.73 Å) than the corresponding Mo-S* one (2.20 Å). This is because oxygen is

much smaller than the sulfido-group and on the other hand has a higher electronegativity which leads to a tighter binding (shorter bond) with the metal. This leads on the other hand to an electronic repulsion of the other ligands (dithiolene, Cys) of the metal which might cause the increase in the respective bonds. An exception is the S(7)-Mo-bond being smaller in the MoVI_O* model than in the MoVI_S* model. This enlarged first coordination sphere of the MoVI_O* model may also be due to the different MO-structure: Since oxygen does not have any d-orbitals, in contrast to the sulfido-ligand, a different arrangement of the MO's might be the reason for the differences in structure. Binding-angles as depicted in Table G1 are very similar in both models. Whether these structural differences may be the reason for the inactivity of the enzyme with oxo-ligated Moco in the prestate has to be further investigated. Especially charge- and electronic-environments, as well as MO-structures have to be studied further for both oxo- and sulfido-ligations at the Moco-site.

Second, the introduction of a hydrogen at the oxo- or sulfido-ligand at MoVI induces very distinct conformational changes in the model- (Moco-) structures, as depicted in Figure E1. These changes in the first coordination sphere of MoVI lead to an reorientation of the pterins. In MoVI_SH they are more in plane with each other and perpendicular to the sulfido-group, yielding a planar-pyramidal-like structure of the Moco. This is in contrast to the MoVI_S* model, in which the pterin-units are more bend compared to the Mo-dithiolene plane and almost parallel to the orientation of the sulfido-group which leads to a trigonal-prismatic coordination of the metal.

In MoVI_SH the Mo(59)-S*(60) bond is about 2.38 Å, while in MoVI_S* it is only 2.20 Å. This indicates the higher single-bond character of the Mo-S*-bond in the SH-model, in contrast to the more double-bond character in the model with sulfido-ligation at MoVI. Because of this longer Mo-S* bond in MoVI_SH (thus less electronic and steric hindrance at the metal-site) all other ligand-metal bonds (with the exception of the Cys(4)-Mo(59) bond) are getting smaller compared to MoVI_S*. This contractions of the ligand-metal bonds might facilitate bigger changes in the ligand-metal-ligand angles and thus conformational rearrangement of the Moco in MoVI_SH compared to MoVI_S*. In more detail, the SH-group is more tilted (than the S*-ligand), with respect to S(7), with an increased S(7)-Mo(59)-S*(60) angle of about 114.98° in contrast to the one in the MoVI_S* model (80.80°). Further, decreased dithiolene-Mo angles might facilitate a more planar arrangement of the pterins in MoVI_SH (compared to MoVI_S*). SH-ligation leads to a decrease of S(7)-Mo(59)-S(34) of about ~35° compared to the MoVI_S* model. Simultaneously, S(8)-Mo(59)-S(33) gets ~4° bigger. Additionally, the angle between (Cys)S(4)-Mo(59)-S(7) decreases with SH-ligation to 95.5° in comparison to S*-ligation (104.67°) of about ~10°, due to the orientation changes of pterin and displacements of Cys. In MoVI_SH, the angle (Cys)S(4)-Mo(59)-S*(60) is ~18° smaller than the corresponding angle in the MoVI_S* model due to a shift of Cys with respect to its position in the latter model. The Mo(59)-(Cys)S(4) bond is also increased to 2.58 Å (compared to 2.41 Å in MoVI_S*). This means that Cys is not as strongly bound to MoVI in the presence of the SH-ligand. Therefore, while the MoVI_S* model is representing a true hexacoordinated arrangement of the six ligands around MoVI, the introduction of a SH-ligation leads to a more pentacoordinated form with a quadratic pyramidal coordination (s. Figure G1). A similar arrangement of the ligands can also been found for all pentacoordinated MoIV model 1 compounds, indicating that the SH-coordination at MoVI leads to a MoIV-like structure of the Moco. The transfer of hydrogen to the sulfido-group at MoVI might therefore trigger the rearrangement of the Moco-structure which might play an important role in preparing the Moco for reduction (see below for further discussion of possible influence of these structural findings on mechanistic properties).

The MoIV models do not differ in their conformation upon exchange of S*- to SH-liagtion. This is why no structural data of these models is presented here. Similar results are obtained for comparing the MoVI_O* and MoVI_OH structures which are not discussed in further detail here.

Comparison to the homology model [49] of the oxidized *Rc*FDH with present sulfido-ligation at the Moco-site, revealed a good agreement of the bond-lengths of the ligand-metal bonds for the MoVI_S* model. In contrast to this, the investigated angles in the first coordination sphere are in parts in good agreement to MoVI_S*, as well as to MoVI_SH. Contrary to this, visual inspection of the homology model revealed that the Moco therein shows a similar trigonal prismatic conformation as observed in the MoVI_S* model. Therefore, the assumed (as discussed in this work; see section 12. and 13.6.) "native" state for the WT with a Moco with MoVI and bound Cys, as well as ligated S*-group is in a more bend form of the cofactor.

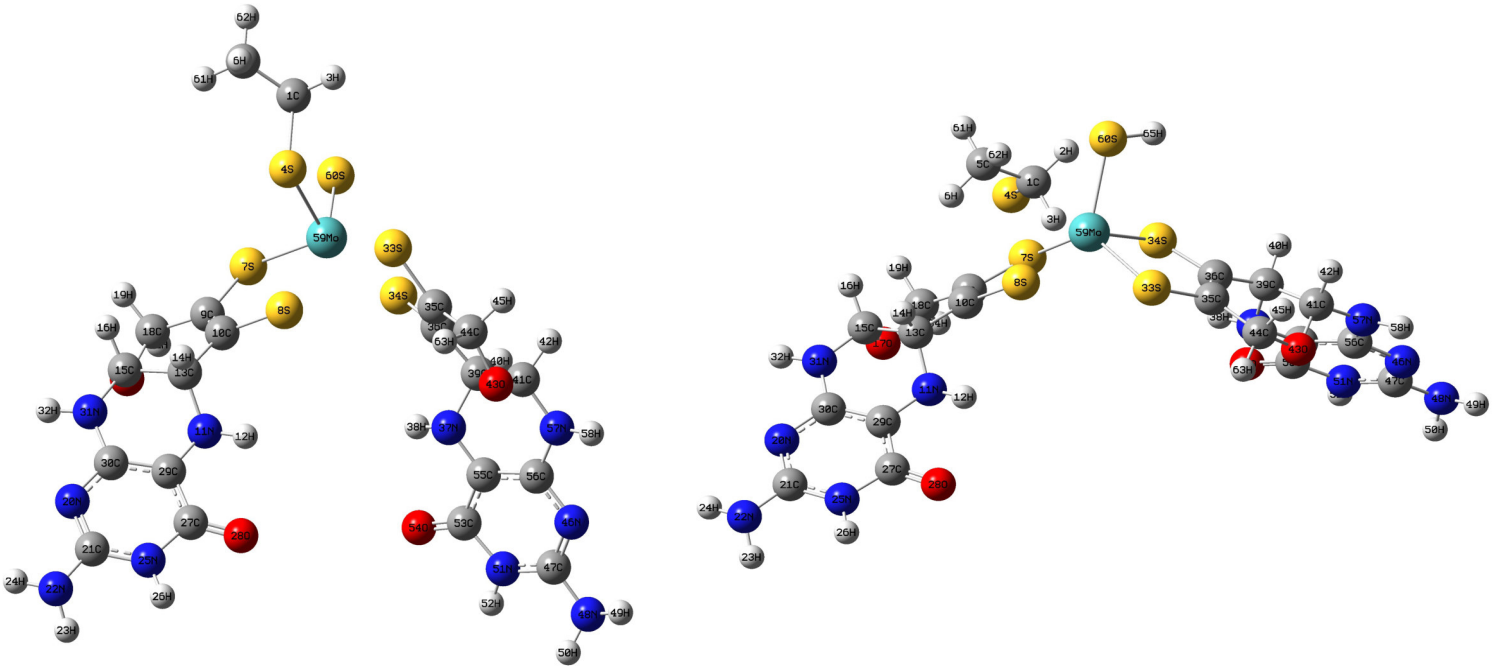

Figure G1. Comparison of QM-optimized structures of hexacoordinated MoVI model 1 compounds with Cys-ligation and S*- or SH-ligation: (left): MoVI_S*; "closed" form; (right): MoVI_SH; "open" form; both after QM-geometry optimizations with ε=4; atoms colored according to type and numbering according to Gaussian [153]

In conclusion, MoVI_S* displays a hexacoordinated coordination of the metal, with bend pterin-units. These are "parallel" to the S*-ligand-orientation. Therefore, the Moco-structure will be considered as "closed". Such a folding-angle has already been observed for high oxidation-states [51],[205]. In contrast to this, the MoVI_SH displays a more pentacoordinated coordination of MoVI with partly dissociated Cys and more planar pterin-units. These are more perpendicular to the SH-ligand. Therefore, this form will be considered as an "open" form. Since the latter form can also be observed for the reduced pentacoordinated MoIV_S*/SH models, these findings indicate a strong relation between structure of the Moco and the reaction cycle and oxidation state of the metal. It may be possible that in addition to the proposed mechanism in section 13.6. not only the ligands (Cys, S*, SH) at the metal-site, but also the structure of the Moco and especially the pterins might play an important role in the oxidation of formate (see Figure G2 for an extended version of the reaction mechanism presented in section 13.6.). It could be that the conformational changes induced by the presence of a SH-ligand at the Moco-site, as found here, might induce conformational strain in the cofactor-binding pocket which initiates a fast reduction and back-oxidation of the Moco, so that it returns from a quadratic pyramidal conformation to a trigonal prismatic form which is better

suited (more stable) at the native active site. This way the strain in the binding pocket would be reduced.

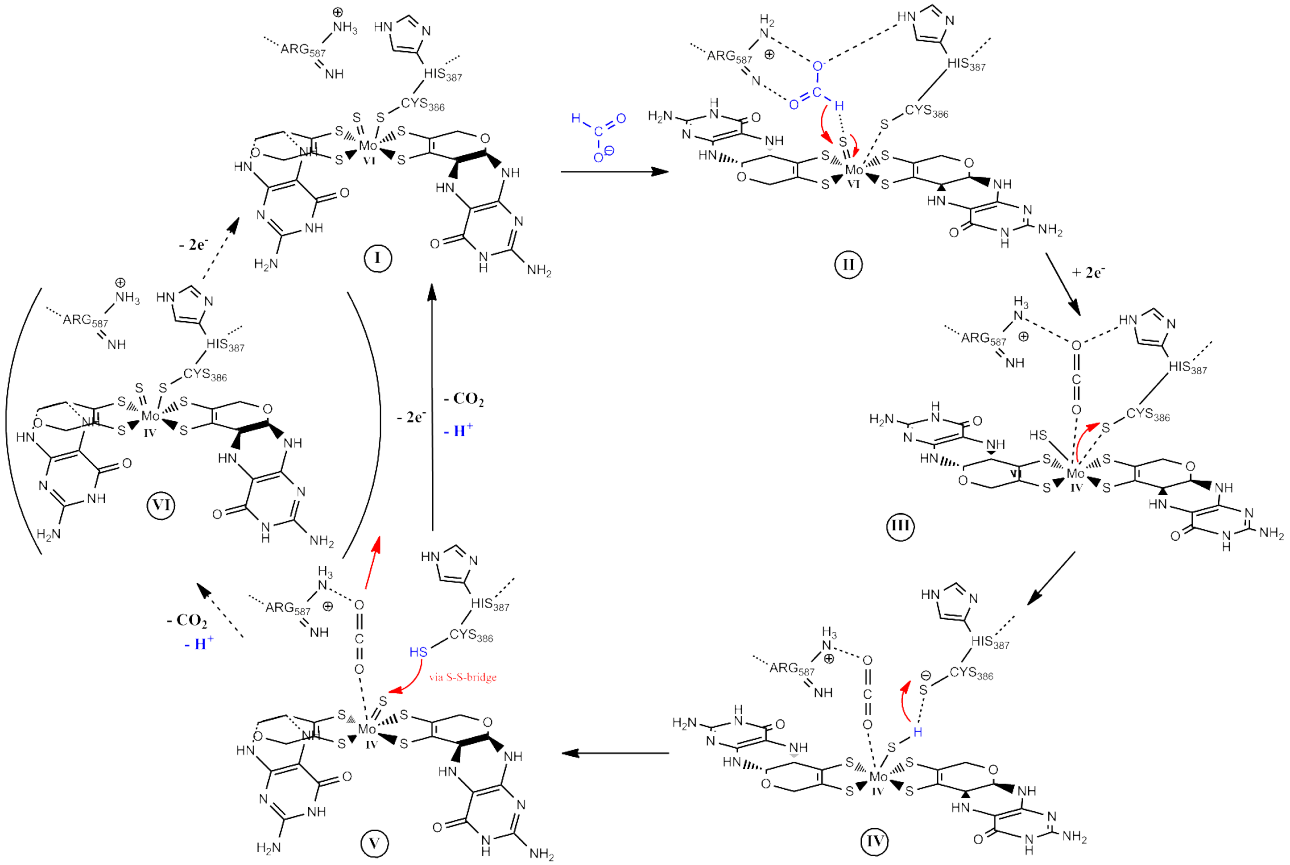

Figure G2. Proposed "Scissors-mechanism" for reversible formate oxidation at the Moco-site of *Rc*FDH using available experimental [73] (Arg587 and His387-interactions with substrate and Moco are due to findings from Hartmann et al. [49] and Maia et al. [54]) and theoretical data and including formal hydride transfer from formate to sulfido-ligand, disclocation of Cys^- from the metal and sulfur-shift mechanistic parts, as well as possible involvement of ligation-sphere-conformations of molybdenum in electron-uptake and -release (orientation of opternes and dithiolenes is simplified to more obviously depict the discussed shift between "closed" (pterins pointing down) and "open" (pterins perpendicular to S*-/SH-ligands) forms of the Moco during the catalytic cycle; please consult text for detailed information about conformational changes)

The reaction cycle starts with a "closed" hexacoordinated Moco (MoVI-) form, with present Cys- and S*-ligations **(I).** Upon approach of formate the hydrogen of $HCOO^-$ is interacting with the sulfido-group yielding partly a SH-bond (**II**). This induces conformational changes in the Moco resulting in a more "open" structure and partly dissociated Cys-moiety (**II**). This prepares the Moco for dissociation of Cys^- and reduction of the metal. After transfer of a formal hydride from formate to the S*-ligand, formation of a SH-ligation and reduction of MoVI to MoIV (**III**), complete "opening" of the Moco,(meaning almost planar pterins) takes place (**III**). Furthermore, Cys is displaced from the metal which yields a pentacoordinated MoIV-form of the Moco (**IV**). Possibly, CO_2 is interacting with the free coordination-site at MoIV.

The "open" form of it may be more appropriate for electron-uptake, since the pterin-units are not redox-innocent and can participate in the reduction process [43],[51] (s. section 2.2. for further details on pterin and dithiolene redox chemistry). They may be important for electron-dislocation inside the molecule and thus stabilizing the reduced cofactor. A more planar arrangement of them

might facilitate this, since planarity is a prerequisite to form a delocalized π-electron system. This possibly makes the "open" conformation of the Moco a better electron-acceptor than the more "closed" form.

Afterward, a proton is abstracted from the SH-group by the nearby unbound Cys^- which induces/prepares the Moco for the formation of the "closed" form of the Moco with almost parallel pterins to the sulfido-ligand-orientation (**V**). This might also be facilitated by the conformational strain between Moco and cofactor binding-pocket, induced by the "open" form of the Moco. The "closed" conformation might be energetically more feasible, since it might fit better into the protein pocket. At least it is more similar to the native structure of the oxidized FDH of *E. coli*. These conformational changes and the free coordination position at MoIV prepare the cofactor for the return of Cys^-. After possible ligation of Cys to MoIV (trigonal prismatic coordination; **VI**), electrons can then be released from the Moco further into the protein matrix. The "closed" form may be a good electron donor since it is a less good electron carrier (as compared to the "open" form) and the pterin ligands can direct the transfer of the electrons. Each pterin may translocate one electron along its orientation in the opposite direction of the sulfido-binding site and into the protein. Finally, the Moco completely returns to its hexacoordinated "closed" form of MoVI (**I**) and the cycle can start again. State **VI** is maybe omitted and the Moco goes directly from **V** to **I**, since Cys-binding to MoIV is energetically less favorable, although Raman-spectra computations suggest the possibility of a hexacoordinated all-sulfur ligation of MoIV (s. 13.3.1. and 13.5.).

Since the Moco changes between a "closed" and an "open" conformation, behaving like a pair of scissors, this new mechanism including the structural influence of the pterin-orientation on the catalytic oxidation of formate, is thus called "Scissor-mechanism".

To verify the influence of the conformation of the molybdenum-containing cofactor on the reaction cycle further studies have to be carried out. The most important would be the determination of the correct crystal structure of *Rc*FDH. Furthermore, vibrational spectroscopy of different model complexes with different sizes, different orientations of pterins and different redox-states of pterin and subsequent comparison to theory may reveal the presence and influence of "open" and "closed" forms. From a theoretical point of view further calculations of models with different pterin redox-states have to be carried out. Also inclusion of additional protein residues into the QM-computations might be important. QM/MM-calculations might provide the best tool for monitoring the influence of different Moco-conformations on the protein environment and *vice versa*. In such a framework the interaction between the Moco in different oxidation states and surrounding protein residues could be studied. Subsequent Raman-spectra computations and comparison to experiments, in which different artificial Mocos are implemented into the native Apo-protein of *Rc*FDH, might shed further light on the problem. Finally, the influence of these reported structural differences in the cofactor on a possible MoV-species has to be investigated accordingly.

The "Scissor-mechanism" provides an extension of the understanding of the catalytic oxidation of formate and might be an example for how not only electronic interactions of individual ligands, but also the complete conformation of the cofactor, might influence the reaction cycle. It is a valuable contribution and expansion to the in 13.6./13.7. and Appendix F presented mechanisms.

H. Mulliken charges computed for (phytochrome Cph1) chromophore PCB and surrounding residues in 4 Å radius centered at PCB and comparison to MM-forcefield charges:

SEGMENT	RES-ID	RESIDUE	ATOM NAME	ATOM TYPE	MULLIKEN-CHARGE	MM-CHARGE	Δ(Mulliken-charge – MM-charge)
PROT	15	LEU	N	NH1	-0.47	-0.47	0
PROT	15	LEU	HN	H	0.31	0.31	0
PROT	15	LEU	CA	CT1	0.07	0.07	0
PROT	15	LEU	HA	HB	0.09	0.09	0
PROT	15	LEU	CB	CT2	-0.388241	-0.18	-0.208241
PROT	15	LEU	HB1	HA	0.122307	0.09	0.032307
PROT	15	LEU	HB2	HA	0.149221	0.09	0.059221
PROT	15	LEU	CG	CT1	-0.015509	-0.09	0.074491
PROT	15	LEU	HG	HA	0.124067	0.09	0.034067
PROT	15	LEU	CD1	CT3	-0.345488	-0.27	-0.075488
PROT	15	LEU	HD11	HA	0.117034	0.09	0.027034
PROT	15	LEU	HD12	HA	0.120383	0.09	0.030383
PROT	15	LEU	HD13	HA	0.122005	0.09	0.032005
PROT	15	LEU	CD2	CT3	-0.36751	-0.27	-0.09751
PROT	15	LEU	HD21	HA	0.125341	0.09	0.035341
PROT	15	LEU	HD22	HA	0.114831	0.09	0.024831
PROT	15	LEU	HD23	HA	0.121553	0.09	0.031553
PROT	15	LEU	C	C	0.51	0.51	0
PROT	15	LEU	O	O	-0.51	-0.51	0
PROT	18	LEU	N	NH1	-0.47	-0.47	0
PROT	18	LEU	HN	H	0.31	0.31	0
PROT	18	LEU	CA	CT1	0.07	0.07	0
PROT	18	LEU	HA	HB	0.09	0.09	0
PROT	18	LEU	CB	CT2	-0.374785	-0.18	-0.194785
PROT	18	LEU	HB1	HA	0.123186	0.09	0.033186
PROT	18	LEU	HB2	HA	0.130805	0.09	0.040805
PROT	18	LEU	CG	CT1	-0.004055	-0.09	0.085945
PROT	18	LEU	HG	HA	0.089845	0.09	-0.000155
PROT	18	LEU	CD1	CT3	-0.36725	-0.27	-0.09725
PROT	18	LEU	HD11	HA	0.13186	0.09	0.04186
PROT	18	LEU	HD12	HA	0.128406	0.09	0.038406
PROT	18	LEU	HD13	HA	0.122739	0.09	0.032739
PROT	18	LEU	CD2	CT3	-0.370214	-0.27	-0.100214
PROT	18	LEU	HD21	HA	0.135302	0.09	0.045302
PROT	18	LEU	HD22	HA	0.138103	0.09	0.048103
PROT	18	LEU	HD23	HA	0.116062	0.09	0.026062
PROT	18	LEU	C	C	0.51	0.51	0
PROT	18	LEU	O	O	-0.51	-0.51	0
PROT	20	ILE	N	NH1	-0.47	-0.47	0
PROT	20	ILE	HN	H	0.31	0.31	0
PROT	20	ILE	CA	CT1	0.07	0.07	0
PROT	20	ILE	HA	HB	0.09	0.09	0
PROT	20	ILE	CB	CT1	-0.228937	-0.09	-0.138937
PROT	20	ILE	HB	HA	0.17341	0.09	0.08341
PROT	20	ILE	CG2	CT3	-0.339834	-0.27	-0.069834

PROT	20	ILE	HG21	HA	0.115604	0.09	0.025604
PROT	20	ILE	HG22	HA	0.112072	0.09	0.022072
PROT	20	ILE	HG23	HA	0.119101	0.09	0.029101
PROT	20	ILE	CG1	CT2	-0.188781	-0.18	-0.008781
PROT	20	ILE	HG11	HA	0.095743	0.09	0.005743
PROT	20	ILE	HG12	HA	0.117204	0.09	0.027204
PROT	20	ILE	CD	CT3	-0.366034	-0.27	-0.096034
PROT	20	ILE	HD1	HA	0.147753	0.09	0.057753
PROT	20	ILE	HD2	HA	0.129609	0.09	0.039609
PROT	20	ILE	HD3	HA	0.113092	0.09	0.023092
PROT	20	ILE	C	C	0.51	0.51	0
PROT	20	ILE	O	O	-0.51	-0.51	0
PROT	174	MET	N	NH1	-0.47	-0.47	0
PROT	174	MET	HN	H	0.31	0.31	0
PROT	174	MET	CA	CT1	0.07	0.07	0
PROT	174	MET	HA	HB	0.09	0.09	0
PROT	174	MET	CB	CT2	-0.36257	-0.18	-0.18257
PROT	174	MET	HB1	HA	0.134253	0.09	0.044253
PROT	174	MET	HB2	HA	0.149798	0.09	0.059798
PROT	174	MET	CG	CT2	-0.128999	-0.14	0.011001
PROT	174	MET	HG1	HA	0.12584	0.09	0.03584
PROT	174	MET	HG2	HA	0.114806	0.09	0.024806
PROT	174	MET	SD	S	-0.191064	-0.09	-0.101064
PROT	174	MET	CE	CT3	-0.298544	-0.22	-0.078544
PROT	174	MET	HE1	HA	0.167974	0.09	0.077974
PROT	174	MET	HE2	HA	0.152218	0.09	0.062218
PROT	174	MET	HE3	HA	0.136292	0.09	0.046292
PROT	174	MET	C	C	0.51	0.51	0
PROT	174	MET	O	O	-0.51	-0.51	0
PROT	176	TYR	N	NH1	-0.47	-0.47	0
PROT	176	TYR	HN	H	0.31	0.31	0
PROT	176	TYR	CA	CT1	0.07	0.07	0
PROT	176	TYR	HA	HB	0.09	0.09	0
PROT	176	TYR	CB	CT2	-0.356792	-0.18	-0.176792
PROT	176	TYR	HB1	HA	0.121674	0.09	0.031674
PROT	176	TYR	HB2	HA	0.141088	0.09	0.051088
PROT	176	TYR	CG	CA	-0.009603	0	-0.009603
PROT	176	TYR	CD1	CA	-0.110569	-0.115	0.004431
PROT	176	TYR	HD1	HP	0.15206	0.115	0.03706
PROT	176	TYR	CE1	CA	-0.213272	-0.115	-0.098272
PROT	176	TYR	HE1	HP	0.172424	0.115	0.057424
PROT	176	TYR	CZ	CA	0.294482	0.11	0.184482
PROT	176	TYR	OH	OH1	-0.500482	-0.54	0.039518
PROT	176	TYR	HH	H	0.386268	0.43	-0.043732
PROT	176	TYR	CD2	CA	-0.142718	-0.115	-0.027718
PROT	176	TYR	HD2	HP	0.138948	0.115	0.023948
PROT	176	TYR	CE2	CA	-0.217877	-0.115	-0.102877
PROT	176	TYR	HE2	HP	0.144376	0.115	0.029376
PROT	176	TYR	C	C	0.51	0.51	0
PROT	176	TYR	O	O	-0.51	-0.51	0

PROT	186	VAL	N	NH1	-0.47	-0.47	0
PROT	186	VAL	HN	H	0.31	0.31	0
PROT	186	VAL	CA	CT1	0.07	0.07	0
PROT	186	VAL	HA	HB	0.09	0.09	0
PROT	186	VAL	CB	CT1	-0.161337	-0.09	-0.071337
PROT	186	VAL	HB	HA	0.134577	0.09	0.044577
PROT	186	VAL	CG1	CT3	-0.277259	-0.27	-0.007259
PROT	186	VAL	HG11	HA	0.161629	0.09	0.071629
PROT	186	VAL	HG12	HA	0.103617	0.09	0.013617
PROT	186	VAL	HG13	HA	-0.006913	0.09	-0.096913
PROT	186	VAL	CG2	CT3	-0.336897	-0.27	-0.066897
PROT	186	VAL	HG21	HA	0.112477	0.09	0.022477
PROT	186	VAL	HG22	HA	0.14227	0.09	0.05227
PROT	186	VAL	HG23	HA	0.127832	0.09	0.037832
PROT	186	VAL	C	C	0.51	0.51	0
PROT	186	VAL	O	O	-0.51	-0.51	0
PROT	198	TYR	N	NH1	-0.47	-0.47	0
PROT	198	TYR	HN	H	0.31	0.31	0
PROT	198	TYR	CA	CT1	0.07	0.07	0
PROT	198	TYR	HA	HB	0.09	0.09	0
PROT	198	TYR	CB	CT2	-0.320228	-0.18	-0.140228
PROT	198	TYR	HB1	HA	0.13122	0.09	0.04122
PROT	198	TYR	HB2	HA	0.088158	0.09	-0.001842
PROT	198	TYR	CG	CA	0.07923	0	0.07923
PROT	198	TYR	CD1	CA	-0.041797	-0.115	0.073203
PROT	198	TYR	HD1	HP	0.130273	0.115	0.015273
PROT	198	TYR	CE1	CA	-0.205867	-0.115	-0.090867
PROT	198	TYR	HE1	HP	0.128893	0.115	0.013893
PROT	198	TYR	CZ	CA	0.272749	0.11	0.162749
PROT	198	TYR	OH	OH1	-0.574868	-0.54	-0.034868
PROT	198	TYR	HH	H	0.422086	0.43	-0.007914
PROT	198	TYR	CD2	CA	-0.139456	-0.115	-0.024456
PROT	198	TYR	HD2	HP	0.114472	0.115	-0.000528
PROT	198	TYR	CE2	CA	-0.216323	-0.115	-0.101323
PROT	198	TYR	HE2	HP	0.131456	0.115	0.016456
PROT	198	TYR	C	C	0.51	0.51	0
PROT	198	TYR	O	O	-0.51	-0.51	0
PROT	203	TYR	N	NH1	-0.47	-0.47	0
PROT	203	TYR	HN	H	0.31	0.31	0
PROT	203	TYR	CA	CT1	0.07	0.07	0
PROT	203	TYR	HA	HB	0.09	0.09	0
PROT	203	TYR	CB	CT2	-0.391909	-0.18	-0.211909
PROT	203	TYR	HB1	HA	0.136723	0.09	0.046723
PROT	203	TYR	HB2	HA	0.148289	0.09	0.058289
PROT	203	TYR	CG	CA	0.022504	0	0.022504
PROT	203	TYR	CD1	CA	-0.090878	-0.115	0.024122
PROT	203	TYR	HD1	HP	0.140472	0.115	0.025472
PROT	203	TYR	CE1	CA	-0.182754	-0.115	-0.067754
PROT	203	TYR	HE1	HP	0.163165	0.115	0.048165
PROT	203	TYR	CZ	CA	0.302513	0.11	0.192513

PROT	203	TYR	OH	OH1	-0.423122	-0.54	0.116878
PROT	203	TYR	HH	H	0.388215	0.43	-0.041785
PROT	203	TYR	CD2	CA	-0.205425	-0.115	-0.090425
PROT	203	TYR	HD2	HP	0.043344	0.115	-0.071656
PROT	203	TYR	CE2	CA	-0.20788	-0.115	-0.09288
PROT	203	TYR	HE2	HP	0.156749	0.115	0.041749
PROT	203	TYR	C	C	0.51	0.51	0
PROT	203	TYR	O	O	-0.51	-0.51	0
PROT	206	SER	N	NH1	-0.47	-0.47	0
PROT	206	SER	HN	H	0.31	0.31	0
PROT	206	SER	CA	CT1	0.07	0.07	0
PROT	206	SER	HA	HB	0.09	0.09	0
PROT	206	SER	CB	CT2	-0.071707	0.05	-0.121707
PROT	206	SER	HB1	HA	0.121212	0.09	0.031212
PROT	206	SER	HB2	HA	0.093504	0.09	0.003504
PROT	206	SER	OG	OH1	-0.518759	-0.66	0.141241
PROT	206	SER	HG1	H	0.375752	0.43	-0.054248
PROT	206	SER	C	C	0.51	0.51	0
PROT	206	SER	O	O	-0.51	-0.51	0
PROT	207	ASP	N	NH1	-0.47	-0.47	0
PROT	207	ASP	HN	H	0.31	0.31	0
PROT	207	ASP	CA	CT1	0.07	0.07	0
PROT	207	ASP	HA	HB	0.09	0.09	0
PROT	207	ASP	CB	CT2	-0.487881	-0.28	-0.207881
PROT	207	ASP	HB1	HA	0.128032	0.09	0.038032
PROT	207	ASP	HB2	HA	0.147121	0.09	0.057121
PROT	207	ASP	CG	CC	0.882511	0.62	0.262511
PROT	207	ASP	OD1	OC	-0.769544	-0.76	-0.009544
PROT	207	ASP	OD2	OC	-0.900236	-0.76	-0.140236
PROT	207	ASP	C	C	0.51	0.51	0
PROT	207	ASP	O	O	-0.51	-0.51	0
PROT	208	ILE	N	NH1	-0.47	-0.47	0
PROT	208	ILE	HN	H	0.31	0.31	0
PROT	208	ILE	CA	CT1	0.07	0.07	0
PROT	208	ILE	HA	HB	0.09	0.09	0
PROT	208	ILE	CB	CT1	-0.212647	-0.09	-0.122647
PROT	208	ILE	HB	HA	0.14482	0.09	0.05482
PROT	208	ILE	CG2	CT3	-0.330262	-0.27	-0.060262
PROT	208	ILE	HG21	HA	0.114344	0.09	0.024344
PROT	208	ILE	HG22	HA	0.131001	0.09	0.041001
PROT	208	ILE	HG23	HA	0.11856	0.09	0.02856
PROT	208	ILE	CG1	CT2	-0.177016	-0.18	0.002984
PROT	208	ILE	HG11	HA	0.084081	0.09	-0.005919
PROT	208	ILE	HG12	HA	0.101852	0.09	0.011852
PROT	208	ILE	CD	CT3	-0.339525	-0.27	-0.069525
PROT	208	ILE	HD1	HA	0.133767	0.09	0.043767
PROT	208	ILE	HD2	HA	0.117988	0.09	0.027988
PROT	208	ILE	HD3	HA	0.113039	0.09	0.023039
PROT	208	ILE	C	C	0.498252	0.51	-0.011748
PROT	208	ILE	O	O	-0.636253	-0.51	-0.126253

PROT	209	PRO	N	N	-0.308416	-0.29	-0.018416
PROT	209	PRO	CD	CP3	-0.002854	0	-0.002854
PROT	209	PRO	HD1	HA	0.130809	0.09	0.040809
PROT	209	PRO	HD2	HA	0.129658	0.09	0.039658
PROT	209	PRO	CA	CP1	-0.025506	0.02	-0.045506
PROT	209	PRO	HA	HB	0.132107	0.09	0.042107
PROT	209	PRO	CB	CP2	-0.203438	-0.18	-0.023438
PROT	209	PRO	HB1	HA	0.124916	0.09	0.034916
PROT	209	PRO	HB2	HA	0.116302	0.09	0.026302
PROT	209	PRO	CG	CP2	-0.186872	-0.18	-0.006872
PROT	209	PRO	HG1	HA	0.127729	0.09	0.037729
PROT	209	PRO	HG2	HA	0.10357	0.09	0.01357
PROT	209	PRO	C	C	0.51	0.51	0
PROT	209	PRO	O	O	-0.51	-0.51	0
PROT	212	ALA	N	NH1	-0.47	-0.47	0
PROT	212	ALA	HN	H	0.31	0.31	0
PROT	212	ALA	CA	CT1	0.07	0.07	0
PROT	212	ALA	HA	HB	0.09	0.09	0
PROT	212	ALA	CB	CT3	-0.506322	-0.27	-0.236322
PROT	212	ALA	HB1	HA	0.181224	0.09	0.091224
PROT	212	ALA	HB2	HA	0.179318	0.09	0.089318
PROT	212	ALA	HB3	HA	0.145782	0.09	0.055782
PROT	212	ALA	C	C	0.51	0.51	0
PROT	212	ALA	O	O	-0.51	-0.51	0
PROT	215	LEU	N	NH1	-0.47	-0.47	0
PROT	215	LEU	HN	H	0.31	0.31	0
PROT	215	LEU	CA	CT1	0.07	0.07	0
PROT	215	LEU	HA	HB	0.09	0.09	0
PROT	215	LEU	CB	CT2	-0.40338	-0.18	-0.22338
PROT	215	LEU	HB1	HA	0.142578	0.09	0.052578
PROT	215	LEU	HB2	HA	0.14998	0.09	0.05998
PROT	215	LEU	CG	CT1	-0.009433	-0.09	0.080567
PROT	215	LEU	HG	HA	0.10633	0.09	0.01633
PROT	215	LEU	CD1	CT3	-0.353522	-0.27	-0.083522
PROT	215	LEU	HD11	HA	0.124445	0.09	0.034445
PROT	215	LEU	HD12	HA	0.113805	0.09	0.023805
PROT	215	LEU	HD13	HA	0.11592	0.09	0.02592
PROT	215	LEU	CD2	CT3	-0.352056	-0.27	-0.082056
PROT	215	LEU	HD21	HA	0.13616	0.09	0.04616
PROT	215	LEU	HD22	HA	0.129668	0.09	0.039668
PROT	215	LEU	HD23	HA	0.099505	0.09	0.009505
PROT	215	LEU	C	C	0.51	0.51	0
PROT	215	LEU	O	O	-0.51	-0.51	0
PROT	216	PHE	N	NH1	-0.47	-0.47	0
PROT	216	PHE	HN	H	0.31	0.31	0
PROT	216	PHE	CA	CT1	0.07	0.07	0
PROT	216	PHE	HA	HB	0.09	0.09	0
PROT	216	PHE	CB	CT2	-0.353495	-0.18	-0.173495
PROT	216	PHE	HB1	HA	0.135133	0.09	0.045133
PROT	216	PHE	HB2	HA	0.123354	0.09	0.033354

PROT	216	PHE	CG	CA	0.048435	0	0.048435
PROT	216	PHE	CD1	CA	-0.142886	-0.115	-0.027886
PROT	216	PHE	HD1	HP	0.16149	0.115	0.04649
PROT	216	PHE	CE1	CA	-0.099953	-0.115	0.015047
PROT	216	PHE	HE1	HP	0.142351	0.115	0.027351
PROT	216	PHE	CZ	CA	-0.115159	-0.115	-0.000159
PROT	216	PHE	HZ	HP	0.129653	0.115	0.014653
PROT	216	PHE	CD2	CA	-0.150527	-0.115	-0.035527
PROT	216	PHE	HD2	HP	0.117128	0.115	0.002128
PROT	216	PHE	CE2	CA	-0.104967	-0.115	0.010033
PROT	216	PHE	HE2	HP	0.109443	0.115	-0.005557
PROT	216	PHE	C	C	0.51	0.51	0
PROT	216	PHE	O	O	-0.51	-0.51	0
PROT	222	ARG	N	NH1	-0.47	-0.47	0
PROT	222	ARG	HN	H	0.31	0.31	0
PROT	222	ARG	CA	CT1	0.07	0.07	0
PROT	222	ARG	HA	HB	0.09	0.09	0
PROT	222	ARG	CB	CT2	-0.350098	-0.18	-0.170098
PROT	222	ARG	HB1	HA	0.156486	0.09	0.066486
PROT	222	ARG	HB2	HA	0.106101	0.09	0.016101
PROT	222	ARG	CG	CT2	-0.17835	-0.18	0.00165
PROT	222	ARG	HG1	HA	0.107412	0.09	0.017412
PROT	222	ARG	HG2	HA	0.118176	0.09	0.028176
PROT	222	ARG	CD	CT2	0.010084	0.2	-0.189916
PROT	222	ARG	HD1	HA	0.077724	0.09	-0.012276
PROT	222	ARG	HD2	HA	0.113248	0.09	0.023248
PROT	222	ARG	NE	NC2	-0.45666	-0.7	0.24334
PROT	222	ARG	HE	HC	0.38919	0.44	-0.05081
PROT	222	ARG	CZ	C	0.710608	0.64	0.070608
PROT	222	ARG	NH1	NC2	-0.650248	-0.8	0.149752
PROT	222	ARG	HH11	HC	0.351915	0.46	-0.108085
PROT	222	ARG	HH12	HC	0.357386	0.46	-0.102614
PROT	222	ARG	NH2	NC2	-0.59061	-0.8	0.20939
PROT	222	ARG	HH21	HC	0.378017	0.46	-0.081983
PROT	222	ARG	HH22	HC	0.349609	0.46	-0.110391
PROT	222	ARG	C	C	0.51	0.51	0
PROT	222	ARG	O	O	-0.51	-0.51	0
PROT	224	ILE	N	NH1	-0.47	-0.47	0
PROT	224	ILE	HN	H	0.31	0.31	0
PROT	224	ILE	CA	CT1	0.07	0.07	0
PROT	224	ILE	HA	HB	0.09	0.09	0
PROT	224	ILE	CB	CT1	-0.190412	-0.09	-0.100412
PROT	224	ILE	HB	HA	0.08699	0.09	-0.00301
PROT	224	ILE	CG2	CT3	-0.3678	-0.27	-0.0978
PROT	224	ILE	HG21	HA	0.10828	0.09	0.01828
PROT	224	ILE	HG22	HA	0.117713	0.09	0.027713
PROT	224	ILE	HG23	HA	0.147928	0.09	0.057928
PROT	224	ILE	CG1	CT2	-0.186459	-0.18	-0.006459
PROT	224	ILE	HG11	HA	0.180616	0.09	0.090616
PROT	224	ILE	HG12	HA	0.118758	0.09	0.028758

PROT	224	ILE	CD	CT3	-0.35701	-0.27	-0.08701
PROT	224	ILE	HD1	HA	0.100891	0.09	0.010891
PROT	224	ILE	HD2	HA	0.111377	0.09	0.021377
PROT	224	ILE	HD3	HA	0.12913	0.09	0.03913
PROT	224	ILE	C	C	0.51	0.51	0
PROT	224	ILE	O	O	-0.51	-0.51	0
PROT	254	ARG	N	NH1	-0.47	-0.47	0
PROT	254	ARG	HN	H	0.31	0.31	0
PROT	254	ARG	CA	CT1	0.07	0.07	0
PROT	254	ARG	HA	HB	0.09	0.09	0
PROT	254	ARG	CB	CT2	-0.374831	-0.18	-0.194831
PROT	254	ARG	HB1	HA	0.133358	0.09	0.043358
PROT	254	ARG	HB2	HA	0.12472	0.09	0.03472
PROT	254	ARG	CG	CT2	-0.229728	-0.18	-0.049728
PROT	254	ARG	HG1	HA	0.139981	0.09	0.049981
PROT	254	ARG	HG2	HA	0.132859	0.09	0.042859
PROT	254	ARG	CD	CT2	-0.02446	0.2	-0.22446
PROT	254	ARG	HD1	HA	0.119117	0.09	0.029117
PROT	254	ARG	HD2	HA	0.164109	0.09	0.074109
PROT	254	ARG	NE	NC2	-0.488346	-0.7	0.211654
PROT	254	ARG	HE	HC	0.358764	0.44	-0.081236
PROT	254	ARG	CZ	C	0.666998	0.64	0.026998
PROT	254	ARG	NH1	NC2	-0.614398	-0.8	0.185602
PROT	254	ARG	HH11	HC	0.357921	0.46	-0.102079
PROT	254	ARG	HH12	HC	0.455092	0.46	-0.004908
PROT	254	ARG	NH2	NC2	-0.656681	-0.8	0.143319
PROT	254	ARG	HH21	HC	0.350921	0.46	-0.109079
PROT	254	ARG	HH22	HC	0.384597	0.46	-0.075403
PROT	254	ARG	C	C	0.51	0.51	0
PROT	254	ARG	O	O	-0.51	-0.51	0
PROT	256	ALA	N	NH1	-0.47	-0.47	0
PROT	256	ALA	HN	H	0.31	0.31	0
PROT	256	ALA	CA	CT1	0.07	0.07	0
PROT	256	ALA	HA	HB	0.09	0.09	0
PROT	256	ALA	CB	CT3	-0.498755	-0.27	-0.228755
PROT	256	ALA	HB1	HA	0.148169	0.09	0.058169
PROT	256	ALA	HB2	HA	0.187486	0.09	0.097486
PROT	256	ALA	HB3	HA	0.1631	0.09	0.0731
PROT	256	ALA	C	C	0.51	0.51	0
PROT	256	ALA	O	O	-0.51	-0.51	0
PROT	257	TYR	N	NH1	-0.47	-0.47	0
PROT	257	TYR	HN	H	0.31	0.31	0
PROT	257	TYR	CA	CT1	0.07	0.07	0
PROT	257	TYR	HA	HB	0.09	0.09	0
PROT	257	TYR	CB	CT2	-0.35264	-0.18	-0.17264
PROT	257	TYR	HB1	HA	0.108676	0.09	0.018676
PROT	257	TYR	HB2	HA	0.169388	0.09	0.079388
PROT	257	TYR	CG	CA	0.026545	0	0.026545
PROT	257	TYR	CD1	CA	-0.119872	-0.115	-0.004872
PROT	257	TYR	HD1	HP	0.133936	0.115	0.018936

PROT	257	TYR	CE1	CA	-0.192589	-0.115	-0.077589
PROT	257	TYR	HE1	HP	0.141742	0.115	0.026742
PROT	257	TYR	CZ	CA	0.302784	0.11	0.192784
PROT	257	TYR	OH	OH1	-0.526649	-0.54	0.013351
PROT	257	TYR	HH	H	0.396668	0.43	-0.033332
PROT	257	TYR	CD2	CA	-0.144779	-0.115	-0.029779
PROT	257	TYR	HD2	HP	0.122165	0.115	0.007165
PROT	257	TYR	CE2	CA	-0.213002	-0.115	-0.098002
PROT	257	TYR	HE2	HP	0.147631	0.115	0.032631
PROT	257	TYR	C	C	0.51	0.51	0
PROT	257	TYR	O	O	-0.51	-0.51	0
PROT	259	CYS	N	NH1	-0.47	-0.47	0
PROT	259	CYS	HN	H	0.31	0.31	0
PROT	259	CYS	CA	CT1	0.07	0.07	0
PROT	259	CYS	HA	HB	0.09	0.09	0
PROT	259	CYS	CB	CT2	-0.261723	-0.1	-0.161723
PROT	259	CYS	HB1	HA	0.110236	0.09	0.020236
PROT	259	CYS	HB2	HA	0.122625	0.09	0.032625
PROT	259	CYS	SG	SE	-0.223663	-0.08	-0.143663
PROT	259	CYS	C	C	0.51	0.51	0
PROT	259	CYS	O	O	-0.51	-0.51	0
PROT	260	HSE	N	NH1	-0.47	-0.47	0
PROT	260	HSE	HN	H	0.31	0.31	0
PROT	260	HSE	CA	CT1	0.07	0.07	0
PROT	260	HSE	HA	HB	0.09	0.09	0
PROT	260	HSE	CB	CT2	-0.322416	-0.08	-0.242416
PROT	260	HSE	HB1	HA	0.137465	0.09	0.047465
PROT	260	HSE	HB2	HA	0.10887	0.09	0.01887
PROT	260	HSE	ND1	NR2	-0.664907	-0.7	0.035093
PROT	260	HSE	CG	CPH1	0.223752	0.22	0.003752
PROT	260	HSE	CE1	CPH2	0.255943	0.25	0.005943
PROT	260	HSE	HE1	HR1	0.108289	0.13	-0.021711
PROT	260	HSE	NE2	NR1	-0.284237	-0.36	0.075763
PROT	260	HSE	HE2	H	0.338536	0.32	0.018536
PROT	260	HSE	CD2	CPH1	-0.059837	-0.05	-0.009837
PROT	260	HSE	HD2	HR3	0.158543	0.09	0.068543
PROT	260	HSE	C	C	0.51	0.51	0
PROT	260	HSE	O	O	-0.51	-0.51	0
PROT	263	TYR	N	NH1	-0.47	-0.47	0
PROT	263	TYR	HN	H	0.31	0.31	0
PROT	263	TYR	CA	CT1	0.07	0.07	0
PROT	263	TYR	HA	HB	0.09	0.09	0
PROT	263	TYR	CB	CT2	-0.364372	-0.18	-0.184372
PROT	263	TYR	HB1	HA	0.152405	0.09	0.062405
PROT	263	TYR	HB2	HA	0.139075	0.09	0.049075
PROT	263	TYR	CG	CA	-0.00603	0	-0.00603
PROT	263	TYR	CD1	CA	-0.143098	-0.115	-0.028098
PROT	263	TYR	HD1	HP	0.138371	0.115	0.023371
PROT	263	TYR	CE1	CA	-0.227682	-0.115	-0.112682
PROT	263	TYR	HE1	HP	0.155093	0.115	0.040093

PROT	263	TYR	CZ	CA	0.282749	0.11	0.172749
PROT	263	TYR	OH	OH1	-0.473881	-0.54	0.066119
PROT	263	TYR	HH	H	0.385718	0.43	-0.044282
PROT	263	TYR	CD2	CA	-0.106976	-0.115	0.008024
PROT	263	TYR	HD2	HP	0.143897	0.115	0.028897
PROT	263	TYR	CE2	CA	-0.223895	-0.115	-0.108895
PROT	263	TYR	HE2	HP	0.14862	0.115	0.03362
PROT	263	TYR	C	C	0.51	0.51	0
PROT	263	TYR	O	O	-0.51	-0.51	0
PROT	267	MET	N	NH1	-0.47	-0.47	0
PROT	267	MET	HN	H	0.31	0.31	0
PROT	267	MET	CA	CT1	0.07	0.07	0
PROT	267	MET	HA	HB	0.09	0.09	0
PROT	267	MET	CB	CT2	-0.356292	-0.18	-0.176292
PROT	267	MET	HB1	HA	0.133414	0.09	0.043414
PROT	267	MET	HB2	HA	0.145326	0.09	0.055326
PROT	267	MET	CG	CT2	-0.11565	-0.14	0.02435
PROT	267	MET	HG1	HA	0.118077	0.09	0.028077
PROT	267	MET	HG2	HA	0.123286	0.09	0.033286
PROT	267	MET	SD	S	-0.175722	-0.09	-0.085722
PROT	267	MET	CE	CT3	-0.271589	-0.22	-0.051589
PROT	267	MET	HE1	HA	0.141564	0.09	0.051564
PROT	267	MET	HE2	HA	0.125906	0.09	0.035906
PROT	267	MET	HE3	HA	0.131686	0.09	0.041686
PROT	267	MET	C	C	0.51	0.51	0
PROT	267	MET	O	O	-0.51	-0.51	0
PROT	272	SER	N	NH1	-0.47	-0.47	0
PROT	272	SER	HN	H	0.31	0.31	0
PROT	272	SER	CA	CT1	0.07	0.07	0
PROT	272	SER	HA	HB	0.09	0.09	0
PROT	272	SER	CB	CT2	-0.053217	0.05	-0.103217
PROT	272	SER	HB1	HA	0.113333	0.09	0.023333
PROT	272	SER	HB2	HA	0.086326	0.09	-0.003674
PROT	272	SER	OG	OH1	-0.603572	-0.66	0.056428
PROT	272	SER	HG1	H	0.457128	0.43	0.027128
PROT	272	SER	C	C	0.51	0.51	0
PROT	272	SER	O	O	-0.51	-0.51	0
PROT	274	THR	N	NH1	-0.47	-0.47	0
PROT	274	THR	HN	H	0.31	0.31	0
PROT	274	THR	CA	CT1	0.07	0.07	0
PROT	274	THR	HA	HB	0.09	0.09	0
PROT	274	THR	CB	CT1	0.054032	0.14	-0.085968
PROT	274	THR	HB	HA	0.108179	0.09	0.018179
PROT	274	THR	OG1	OH1	-0.528298	-0.66	0.131702
PROT	274	THR	HG1	H	0.371569	0.43	-0.058431
PROT	274	THR	CG2	CT3	-0.37791	-0.27	-0.10791
PROT	274	THR	HG21	HA	0.139826	0.09	0.049826
PROT	274	THR	HG22	HA	0.113238	0.09	0.023238
PROT	274	THR	HG23	HA	0.119366	0.09	0.029366
PROT	274	THR	C	C	0.51	0.51	0

PROT	274	THR	O	O	-0.51	-0.51	0
PROT	286	LEU	N	NH1	-0.47	-0.47	0
PROT	286	LEU	HN	H	0.31	0.31	0
PROT	286	LEU	CA	CT1	0.07	0.07	0
PROT	286	LEU	HA	HB	0.09	0.09	0
PROT	286	LEU	CB	CT2	-0.374505	-0.18	-0.194505
PROT	286	LEU	HB1	HA	0.139588	0.09	0.049588
PROT	286	LEU	HB2	HA	0.113361	0.09	0.023361
PROT	286	LEU	CG	CT1	0.005233	-0.09	0.095233
PROT	286	LEU	HG	HA	0.086147	0.09	-0.003853
PROT	286	LEU	CD1	CT3	-0.349518	-0.27	-0.079518
PROT	286	LEU	HD11	HA	0.131199	0.09	0.041199
PROT	286	LEU	HD12	HA	0.110088	0.09	0.020088
PROT	286	LEU	HD13	HA	0.114669	0.09	0.024669
PROT	286	LEU	CD2	CT3	-0.356078	-0.27	-0.086078
PROT	286	LEU	HD21	HA	0.118776	0.09	0.028776
PROT	286	LEU	HD22	HA	0.137305	0.09	0.047305
PROT	286	LEU	HD23	HA	0.12373	0.09	0.03373
PROT	286	LEU	C	C	0.51	0.51	0
PROT	286	LEU	O	O	-0.51	-0.51	0
PROT	288	ALA	N	NH1	-0.47	-0.47	0
PROT	288	ALA	HN	H	0.31	0.31	0
PROT	288	ALA	CA	CT1	0.07	0.07	0
PROT	288	ALA	HA	HB	0.09	0.09	0
PROT	288	ALA	CB	CT3	-0.500196	-0.27	-0.230196
PROT	288	ALA	HB1	HA	0.169684	0.09	0.079684
PROT	288	ALA	HB2	HA	0.167661	0.09	0.077661
PROT	288	ALA	HB3	HA	0.162852	0.09	0.072852
PROT	288	ALA	C	C	0.51	0.51	0
PROT	288	ALA	O	O	-0.51	-0.51	0
PROT	290	HSE	N	NH1	-0.47	-0.47	0
PROT	290	HSE	HN	H	0.31	0.31	0
PROT	290	HSE	CA	CT1	0.07	0.07	0
PROT	290	HSE	HA	HB	0.09	0.09	0
PROT	290	HSE	CB	CT2	-0.345721	-0.08	-0.265721
PROT	290	HSE	HB1	HA	0.117821	0.09	0.027821
PROT	290	HSE	HB2	HA	0.117042	0.09	0.027042
PROT	290	HSE	ND1	NR2	-0.66698	-0.7	0.03302
PROT	290	HSE	CG	CPH1	0.207805	0.22	-0.012195
PROT	290	HSE	CE1	CPH2	0.291802	0.25	0.041802
PROT	290	HSE	HE1	HR1	0.134676	0.13	0.004676
PROT	290	HSE	NE2	NR1	-0.255107	-0.36	0.104893
PROT	290	HSE	HE2	H	0.367557	0.32	0.047557
PROT	290	HSE	CD2	CPH1	-0.115437	-0.05	-0.065437
PROT	290	HSE	HD2	HR3	0.146541	0.09	0.056541
PROT	290	HSE	C	C	0.51	0.51	0
PROT	290	HSE	O	O	-0.51	-0.51	0
PROT	457	ALA	N	NH1	-0.47	-0.47	0
PROT	457	ALA	HN	H	0.31	0.31	0
PROT	457	ALA	CA	CT1	0.07	0.07	0

PROT	457	ALA	HA	HB	0.09	0.09	0
PROT	457	ALA	CB	CT3	-0.495258	-0.27	-0.225258
PROT	457	ALA	HB1	HA	0.153424	0.09	0.063424
PROT	457	ALA	HB2	HA	0.173874	0.09	0.083874
PROT	457	ALA	HB3	HA	0.167958	0.09	0.077958
PROT	457	ALA	C	C	0.51	0.51	0
PROT	457	ALA	O	O	-0.51	-0.51	0
PROT	458	TYR	N	NH1	-0.47	-0.47	0
PROT	458	TYR	HN	H	0.31	0.31	0
PROT	458	TYR	CA	CT1	0.07	0.07	0
PROT	458	TYR	HA	HB	0.09	0.09	0
PROT	458	TYR	CB	CT2	-0.356982	-0.18	-0.176982
PROT	458	TYR	HB1	HA	0.127765	0.09	0.037765
PROT	458	TYR	HB2	HA	0.149629	0.09	0.059629
PROT	458	TYR	CG	CA	-0.001194	0	-0.001194
PROT	458	TYR	CD1	CA	-0.1165	-0.115	-0.0015
PROT	458	TYR	HD1	HP	0.149632	0.115	0.034632
PROT	458	TYR	CE1	CA	-0.242143	-0.115	-0.127143
PROT	458	TYR	HE1	HP	0.154263	0.115	0.039263
PROT	458	TYR	CZ	CA	0.254099	0.11	0.144099
PROT	458	TYR	OH	OH1	-0.471259	-0.54	0.068741
PROT	458	TYR	HH	H	0.383708	0.43	-0.046292
PROT	458	TYR	CD2	CA	-0.110696	-0.115	0.004304
PROT	458	TYR	HD2	HP	0.130113	0.115	0.015113
PROT	458	TYR	CE2	CA	-0.204783	-0.115	-0.089783
PROT	458	TYR	HE2	HP	0.154352	0.115	0.039352
PROT	458	TYR	C	C	0.51	0.51	0
PROT	458	TYR	O	O	-0.51	-0.51	0
PROT	469	LEU	N	NH1	-0.47	-0.47	0
PROT	469	LEU	HN	H	0.31	0.31	0
PROT	469	LEU	CA	CT1	0.07	0.07	0
PROT	469	LEU	HA	HB	0.09	0.09	0
PROT	469	LEU	CB	CT2	-0.38947	-0.18	-0.20947
PROT	469	LEU	HB1	HA	0.144261	0.09	0.054261
PROT	469	LEU	HB2	HA	0.113024	0.09	0.023024
PROT	469	LEU	CG	CT1	-0.009374	-0.09	0.080626
PROT	469	LEU	HG	HA	0.106189	0.09	0.016189
PROT	469	LEU	CD1	CT3	-0.340292	-0.27	-0.070292
PROT	469	LEU	HD11	HA	0.123287	0.09	0.033287
PROT	469	LEU	HD12	HA	0.122671	0.09	0.032671
PROT	469	LEU	HD13	HA	0.109349	0.09	0.019349
PROT	469	LEU	CD2	CT3	-0.3675	-0.27	-0.0975
PROT	469	LEU	HD21	HA	0.116665	0.09	0.026665
PROT	469	LEU	HD22	HA	0.146179	0.09	0.056179
PROT	469	LEU	HD23	HA	0.125007	0.09	0.035007
PROT	469	LEU	C	C	0.51	0.51	0
PROT	469	LEU	O	O	-0.51	-0.51	0
PROT	470	HSD	N	NH1	-0.47	-0.47	0
PROT	470	HSD	HN	H	0.31	0.31	0
PROT	470	HSD	CA	CT1	0.07	0.07	0

PROT	470	HSD	HA	HB	0.09	0.09	0
PROT	470	HSD	CB	CT2	-0.342252	-0.09	-0.252252
PROT	470	HSD	HB1	HA	0.124031	0.09	0.034031
PROT	470	HSD	HB2	HA	0.171916	0.09	0.081916
PROT	470	HSD	ND1	NR1	-0.292259	-0.36	0.067741
PROT	470	HSD	HD1	H	0.344062	0.32	0.024062
PROT	470	HSD	CG	CPH1	0.067409	-0.05	0.117409
PROT	470	HSD	CE1	CPH2	0.275868	0.25	0.025868
PROT	470	HSD	HE1	HR1	0.098607	0.13	-0.031393
PROT	470	HSD	NE2	NR2	-0.605171	-0.7	0.094829
PROT	470	HSD	CD2	CPH1	-0.000318	0.22	-0.220318
PROT	470	HSD	HD2	HR3	0.15811	0.1	0.05811
PROT	470	HSD	C	C	0.485981	0.51	-0.024019
PROT	470	HSD	O	O	-0.629798	-0.51	-0.119798
PROT	471	PRO	N	N	-0.312691	-0.29	-0.022691
PROT	471	PRO	CD	CP3	-0.001116	0	-0.001116
PROT	471	PRO	HD1	HA	0.112821	0.09	0.022821
PROT	471	PRO	HD2	HA	0.123409	0.09	0.033409
PROT	471	PRO	CA	CP1	-0.026605	0.02	-0.046605
PROT	471	PRO	HA	HB	0.123869	0.09	0.033869
PROT	471	PRO	CB	CP2	-0.187684	-0.18	-0.007684
PROT	471	PRO	HB1	HA	0.117254	0.09	0.027254
PROT	471	PRO	HB2	HA	0.125963	0.09	0.035963
PROT	471	PRO	CG	CP2	-0.182719	-0.18	-0.002719
PROT	471	PRO	HG1	HA	0.119297	0.09	0.029297
PROT	471	PRO	HG2	HA	0.132022	0.09	0.042022
PROT	471	PRO	C	C	0.51	0.51	0
PROT	471	PRO	O	O	-0.51	-0.51	0
PCB	515	PCB	CAC	CT1	0.022372	-0.159	0.181372
PCB	515	PCB	HAC	HA	0.124037	0.08	0.044037
PCB	515	PCB	C1C	C	0.546957	0.289	0.257957
PCB	515	PCB	H2C	HA	0.135234	0.057	0.078234
PCB	515	PCB	N_C	NR1	-0.309234	-0.5	0.190766
PCB	515	PCB	H_C	H	0.357852	0.302	0.055852
PCB	515	PCB	C4C	CA	0.219887	0.237	-0.017113
PCB	515	PCB	C3C	CT1	-0.070796	0.017	-0.087796
PCB	515	PCB	H3C	HA	0.131418	0.144	-0.012582
PCB	515	PCB	C2C	CT1	-0.122319	0.365	-0.487319
PCB	515	PCB	O_C	O	-0.620204	-0.495	-0.125204
PCB	515	PCB	CHD	CPY3	-0.276514	-0.501	0.224486
PCB	515	PCB	HHD	HA	0.145539	0.214	-0.068461
PCB	515	PCB	C1D	CPA	0.225633	0.375	-0.149367
PCB	515	PCB	N_D	NR1	-0.310071	-0.543	0.232929
PCB	515	PCB	H_D	H	0.361425	0.306	0.055425
PCB	515	PCB	C4D	CPA	0.203474	0.364	-0.160526
PCB	515	PCB	C3D	CPB	-0.012262	-0.041	0.028738
PCB	515	PCB	C2D	CPB	-0.084337	-0.161	0.076663
PCB	515	PCB	CHA	CPM	-0.113111	-0.111	-0.002111
PCB	515	PCB	HHA	HA	0.148285	0.319	-0.170715
PCB	515	PCB	C1A	CPA	0.123117	0.169	-0.045883

PCB	515	PCB	N_A	NR1	-0.311569	-0.654	0.342431
PCB	515	PCB	H_A	H	0.362276	0.355	0.007276
PCB	515	PCB	C2A	CPY4	-0.038603	-0.038	-0.000603
PCB	515	PCB	C3A	CPB	-0.05365	-0.175	0.12135
PCB	515	PCB	C4A	CPA	0.179525	0.437	-0.257475
PCB	515	PCB	C4B	C	0.555147	0.306	0.249147
PCB	515	PCB	N_B	NR1	-0.384713	-0.56	0.175287
PCB	515	PCB	H_B	H	0.363032	0.375	-0.011968
PCB	515	PCB	C1B	CA	0.198336	0.452	-0.253664
PCB	515	PCB	C2B	CPY5	-0.04878	-0.37	0.32122
PCB	515	PCB	C3B	CPY6	-0.122826	0.32	-0.442826
PCB	515	PCB	O_B	O	-0.798318	-0.337	-0.461318
PCB	515	PCB	CHB	CPY2	-0.212589	-0.448	0.235411
PCB	515	PCB	HHB	HA	0.129106	0.219	-0.089894
PCB	515	PCB	CAB	CT2	-0.14683	-0.18	0.03317
PCB	515	PCB	HAB	HA	0.099661	0.09	0.009661
PCB	515	PCB	HBB	HA	0.10139	0.09	0.01139
PCB	515	PCB	CBB	CT3	-0.35268	-0.27	-0.08268
PCB	515	PCB	HV1	HA	0.140111	0.09	0.050111
PCB	515	PCB	HV2	HA	0.128107	0.09	0.038107
PCB	515	PCB	HV3	HA	0.112763	0.09	0.022763
PCB	515	PCB	CBC	CT3	-0.369827	-0.16	-0.209827
PCB	515	PCB	HL1	HA	0.119615	0.09	0.029615
PCB	515	PCB	HL2	HA	0.120193	0.09	0.030193
PCB	515	PCB	HL3	HA	0.132602	0.09	0.042602
PCB	515	PCB	CMC	CT3	-0.343987	-0.179	-0.164987
PCB	515	PCB	HE1	HA	0.13511	0.038	0.09711
PCB	515	PCB	HE2	HA	0.132873	0.038	0.094873
PCB	515	PCB	HE3	HA	0.118988	0.038	0.080988
PCB	515	PCB	CMD	CT3	-0.283581	0.056	-0.339581
PCB	515	PCB	HD1	HA	0.12177	0.022	0.09977
PCB	515	PCB	HD2	HA	0.086645	0.022	0.064645
PCB	515	PCB	HD3	HA	0.142443	0.022	0.120443
PCB	515	PCB	CMA	CT3	-0.296504	-0.103	-0.193504
PCB	515	PCB	HA1	HA	0.141414	0.066	0.075414
PCB	515	PCB	HA2	HA	0.132355	0.066	0.066355
PCB	515	PCB	HA3	HA	0.085675	0.066	0.019675
PCB	515	PCB	CMB	CT3	-0.339255	-0.018	-0.321255
PCB	515	PCB	HB1	HA	0.154571	0.049	0.105571
PCB	515	PCB	HB2	HA	0.13463	0.049	0.08563
PCB	515	PCB	HB3	HA	0.135575	0.049	0.086575
PCB	515	PCB	CBD	CT2	-0.28582	-0.28	-0.00582
PCB	515	PCB	HO3	HA	0.0999	0.09	0.0099
PCB	515	PCB	HO4	HA	0.106301	0.09	0.016301
PCB	515	PCB	CGD	CC	0.79842	0.62	0.17842
PCB	515	PCB	O2D	OC	-0.797392	-0.76	-0.037392
PCB	515	PCB	O1D	OC	-0.874886	-0.76	-0.114886
PCB	515	PCB	CBA	CT2	-0.340253	-0.28	-0.060253
PCB	515	PCB	HO1	HA	0.135088	0.09	0.045088
PCB	515	PCB	HO2	HA	0.125557	0.09	0.035557

PCB	515	PCB	CGA	CC	0.857892	0.62	0.237892
PCB	515	PCB	O2A	OC	-0.851849	-0.76	-0.091849
PCB	515	PCB	O1A	OC	-0.88939	-0.76	-0.12939
PCB	515	PCB	CAA	CT2	-0.153064	-0.18	0.026936
PCB	515	PCB	HO5	HA	0.141991	0.09	0.051991
PCB	515	PCB	HO6	HA	0.102286	0.09	0.012286
PCB	515	PCB	CAD	CT2	-0.121517	-0.18	0.058483
PCB	515	PCB	HO7	HA	0.100949	0.09	0.010949
PCB	515	PCB	HO8	HA	0.111695	0.09	0.021695
HOH	2	TIP3	OH2	OT	-0.789255	-0.834	0.044745
HOH	2	TIP3	H1	HT	0.381769	0.417	-0.035231
HOH	2	TIP3	H2	HT	0.407486	0.417	-0.009514
HOH	3	TIP3	OH2	OT	-0.788255	-0.834	0.045745
HOH	3	TIP3	H1	HT	0.38685	0.417	-0.03015
HOH	3	TIP3	H2	HT	0.401405	0.417	-0.015595
HOH	4	TIP3	OH2	OT	-0.845133	-0.834	-0.011133
HOH	4	TIP3	H1	HT	0.40822	0.417	-0.00878
HOH	4	TIP3	H2	HT	0.436913	0.417	0.019913
HOH	10	TIP3	OH2	OT	-0.870272	-0.834	-0.036272
HOH	10	TIP3	H1	HT	0.411348	0.417	-0.005652
HOH	10	TIP3	H2	HT	0.458924	0.417	0.041924
HOH	11	TIP3	OH2	OT	-0.81152	-0.834	0.02248
HOH	11	TIP3	H1	HT	0.421003	0.417	0.004003
HOH	11	TIP3	H2	HT	0.390516	0.417	-0.026484
HOH	28	TIP3	OH2	OT	-0.823785	-0.834	0.010215
HOH	28	TIP3	H1	HT	0.431861	0.417	0.014861
HOH	28	TIP3	H2	HT	0.391924	0.417	-0.025076
WT33	632	TIP3	OH2	OT	-0.739864	-0.834	0.094136
WT33	632	TIP3	H1	HT	0.39677	0.417	-0.02023
WT33	632	TIP3	H2	HT	0.343093	0.417	-0.073907
WT33	663	TIP3	OH2	OT	-0.74382	-0.834	0.09018
WT33	663	TIP3	H1	HT	0.354992	0.417	-0.062008
WT33	663	TIP3	H2	HT	0.388828	0.417	-0.028172
WT34	566	TIP3	OH2	OT	-0.758314	-0.834	0.075686
WT34	566	TIP3	H1	HT	0.381755	0.417	-0.035245
WT34	566	TIP3	H2	HT	0.376559	0.417	-0.040441
WT34	572	TIP3	OH2	OT	-0.71291	-0.834	0.12109
WT34	572	TIP3	H1	HT	0.352872	0.417	-0.064128
WT34	572	TIP3	H2	HT	0.360038	0.417	-0.056962
WT34	577	TIP3	OH2	OT	-0.762261	-0.834	0.071739
WT34	577	TIP3	H1	HT	0.399952	0.417	-0.017048
WT34	577	TIP3	H2	HT	0.362308	0.417	-0.054692
WT34	578	TIP3	OH2	OT	-0.757433	-0.834	0.076567
WT34	578	TIP3	H1	HT	0.376292	0.417	-0.040708
WT34	578	TIP3	H2	HT	0.381141	0.417	-0.035859
WT34	579	TIP3	OH2	OT	-0.720379	-0.834	0.113621
WT34	579	TIP3	H1	HT	0.372283	0.417	-0.044717
WT34	579	TIP3	H2	HT	0.348096	0.417	-0.068904
WT34	586	TIP3	OH2	OT	-0.734053	-0.834	0.099947
WT34	586	TIP3	H1	HT	0.376527	0.417	-0.040473

WT34	586	TIP3	H2	HT	0.357526	0.417	-0.059474
WT34	589	TIP3	OH2	OT	-0.7813	-0.834	0.0527
WT34	589	TIP3	H1	HT	0.373596	0.417	-0.043404
WT34	589	TIP3	H2	HT	0.407705	0.417	-0.009295
WT34	605	TIP3	OH2	OT	-0.722373	-0.834	0.111627
WT34	605	TIP3	H1	HT	0.362455	0.417	-0.054545
WT34	605	TIP3	H2	HT	0.359918	0.417	-0.057082
WT34	614	TIP3	OH2	OT	-0.742344	-0.834	0.091656
WT34	614	TIP3	H1	HT	0.380188	0.417	-0.036812
WT34	614	TIP3	H2	HT	0.362156	0.417	-0.054844
WT37	610	TIP3	OH2	OT	-0.718098	-0.834	0.115902
WT37	610	TIP3	H1	HT	0.365566	0.417	-0.051434
WT37	610	TIP3	H2	HT	0.352532	0.417	-0.064468
WT37	619	TIP3	OH2	OT	-0.754945	-0.834	0.079055
WT37	619	TIP3	H1	HT	0.394147	0.417	-0.022853
WT37	619	TIP3	H2	HT	0.360798	0.417	-0.056202
WT37	620	TIP3	OH2	OT	-0.723615	-0.834	0.110385
WT37	620	TIP3	H1	HT	0.352866	0.417	-0.064134
WT37	620	TIP3	H2	HT	0.370749	0.417	-0.046251
WT37	624	TIP3	OH2	OT	-0.772765	-0.834	0.061235
WT37	624	TIP3	H1	HT	0.388395	0.417	-0.028605
WT37	624	TIP3	H2	HT	0.38437	0.417	-0.03263
WT37	626	TIP3	OH2	OT	-0.777557	-0.834	0.056443
WT37	626	TIP3	H1	HT	0.413813	0.417	-0.003187
WT37	626	TIP3	H2	HT	0.363744	0.417	-0.053256
WT37	629	TIP3	OH2	OT	-0.729632	-0.834	0.104368
WT37	629	TIP3	H1	HT	0.37538	0.417	-0.04162
WT37	629	TIP3	H2	HT	0.354252	0.417	-0.062748

Bibliography

[1] Butler, W. L.; Norris, K. H.; Siegelman, H. W.; Hendricks, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **1959**, *45*, 1703

[2] Quail, P. H. *Nat. ReV. Mol. Cell Biol.* **2002**, *3*, 85

[3] Rockwell, N. C.; Su, Y. S.; Lagarias, J. C. *Annu. ReV. Plant Biol.* **2006**, *57*, 837

[4] Bhoo, S.; Davis, S.; Walker, J.; Karniol, B.; Vierstra, R. *Nature* **2001**, *414*, 776

[5] Giraud, E. et al. *J. Biol. Chem.* **2005**, *280*, 37

[6] Hübschmann, T.; Börner, T.; Hartmann, E.; Lamparter, T. *Eur. J. Biochem.* **2001**, *268*, 2055

[7] Karniol, B.; Vierstra, R. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2807

[8] Lamparter, T.; Esteban, B.; Hughes, J. *Eur. J. Biochem.* **2001**, *268*, 4720

[9] Wagner, J. R.; Brunzelle, J. S.; Forest, K. T.; Vierstra, R. D. *Nature* **2005**, *438*, 325

[10] Essen, L. O.; Hughes, J.; Mailliet, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14709

[11] Yang, X.; Kuk, J.; Moffat, K. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14715

[12] Kaminski, S.; Daminelli, G.; Mroginski, M. A. *J. Phys. Chem. B* **2009**, *113*, 945

[13] van Thor, J. J.; Mackeen, M.; Kuprow, I.; Dwek, R. A.; Wormold, M. R. *Biophys. J.* **2006**, *91*, 1811

[14] Hahn, J.; Strauss, H. M.; Schmieder, P. *J. Am. Chem. Soc.* **2008**, *130*, 11170

[15] Berman, H.; Battistuz, T.; Bhat, T.; Bluhm, W.; Bourne, P.; Burkhardt, K.; Feng, Z.; Gilliland, G.; Iype, L.; Jain, S. et al. *Acta Cryst. D* **2002**, *58*, 899

[16] Schmidt, A.; Sauthof, L.; Szczepek, M.; Lopez, M.; Escobar, F.; Qureshi, B.; Michael, N.; Buhrke, D.; Stevens, T.; Kwiatkowski, D.; von Stetten, D.; Mroginski, M.; Krauß, N.; Lamparter, T.; Hildebrandt, P. & Scheerer, P. *Nature Communications* **2018**, *9*, 4912; In this thesis (by Dennis Belger), an early unpublished version of the X-ray structure (with an endo-cyclic double bond of ring A of BV) of the aforementioned crystal structure by Scheerer et al. has been used as initial set of coordinates for computations of Agp2.

[17] Kneip, C. *PhD thesis* **1998**, Duisburg, Germany

[18] Kneip, C.; Hildebrandt, P.; Schlamann, W.; Braslavsky, S.; Schaffner, K. *Biochem.* **1999**, *38*, 15185

[19] Matysik, j.; Hildebrandt, P.; Schlamann, W.; Braslavsky, S.; Schaffner, K. *Biochem.* **1995**, *34*, 10497

[20] Rohmer, T. et al. *J. Phys. Chem. B* **2006**, *110*, 20580

[21] Strauss, H. *PhD thesis* **2005**, Berlin, Germany

[22] Strauss, H.; Hughes, J.; Schmieder, P. *Biochem.* **2005**, *44*, 8244

[23] Escobar, F. Velazquez; Piwowarski, P.; Salewski, J.; Michael, N.; Lopez, M. Fernandez; Rupp, A.; Qureshi, B.

Muhammed; Scheerer, P.; Bartl, F.; Frankenberg-Dinkel, N.; Siebert, F.; Mroginski, M. Andrea; Hildebrandt, P. *Nature Chemistry* **2015**, *7*, 423

[24] Rockwell, N.; Shang, L.; Martin, S.; Lagarias, J. *PNAS* **2009**, *106*, 6123

[25] Shang, L.; Rockwell, N.; Martin, S.; Lagarias, J. *Biochem.* **2010**, *49*, 6070

[26] Hildebrandt et al. **unpublished (at the time of writing this thesis)**

[27] Velazquez-Escobar, F.; Lang, C.; Takiden, A.; Schneider, C.; Balke, J.; Hughes, J.; Alexiev, U.; Hildebrandt, P.; Mroginski, M. *J. Phys. Chem. B* **2017**, *121*, 47

[28] Takiden, A.; Velazquez-Escobar, F.; Dragelj, J.; Woelke, A.; Knapp, E.-W.; Piwowarski, P.; Bart, F.; Hildebrandt, P.; Mroginski, M. *Photochem. And Photobiol.* **2017**, *93*, 713

[29] Song, C.; Psakis, G.; Lang, C.; Mailliet, J.; Gartner, W.; Hughes, J.; Matysik, J. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3842

[30] Song et al. *Biochem.* **2011**, *50*, 10987

[31] Braslavsky, S.; Gartner, W.; Schaffner, K. *Plant Cell Environ.* **1997**, *20*, 700

[32] Borucki, B. et al. *J. Biol. Chem.* **2005**, *280*, 34358

[33] Brock, H. et al. *Biochem.* **1987**, *26*, 1412

[34] Foerstendorf, H.; Benda, C.; Gärtner, W.; Storf, M.; Scheer, H.; Siebert, F. *Biochem.* **2001**, *40*, 14952

[35] Foerstendorf, H.; Lamparter, T.; Hughes, J.; Gärtner, W.; Siebert, F. *Photochem. Photobiol.* **2000**, *71*, 655

[36] Foerstendorf, H.; Mummert, E.; Schäfer, E.; Scheer, H.; Siebert, F. *Biochem.* **1996**, *35*, 10793

[37] Andel, K.; Lagarias, J.; Mathies, R. *Biochem.* **1996**, *35*, 15997

[38] Andel, F.; Murphy, J.; Haas, J.; McDowell, M.; van der Hoef, I.; Lugtenburg, J.; Lagarias, J.; Mathies, R. *Biochem.* **2000**, *39*, 2667

[39] Mitzutani, Y.; Tokutomi, S.; Kitagawa, T. *Biochem.* **1994**, *33*, 153

[40] von Stetten, D. et al. *J. Biol. Chem.* **2007**, *282*, 2116

[41] Kaminski, S. *Phd-Thesis* **2010**, Berlin, Germany

[42] Burgie, E.; Zhang, J.; Vierstra, R. *Structure* **2016**, *24*, 1

[43] Perez-Miller, S.; Hurley, T. *Biochem.* **2003**, *42*, 7100

[44] Liu, Z.; Sun, Y.; Rose, J. et al. *Nature Structural Biology* **1997**, *4*, 317

[45] Thomasson, H.; Edenberg, H.; Crabb, D. et al. *American Journal of Human Genetics* **1991**, *48*, 677

[46] Ferry, J. *FEMS Microbiol. Rev.* **1990**, *7*, 377

[47] Khanguiov, S.; Gladyshev, V.; Dismukes, G.; Stadtman, T. *Biochem.* **1998**, *37*, 3518

[48] Hartmann, T.; Leimkühler, S. *FEBS Journal* **2013**, *280*, 6083

[49] Hartmann, T.; Schrapers, P.; Utesch, T.; Nimtz, M.; Rippers, Y.; Dau, H.; Mroginski, M.; Haumann, M.; Leimkühler, S. *Biochem.* **2016**, *55*, 2381

[50] Schrapers, P.; Hartmann, T.; Kositzki, R.; Dau, H.; Reschke, S., Schulzke, C.; Leimkühler, S.; Haumann, M. *Inorg. Chem.* **2015**, *54*, 3260

[51] Basu et al. *J. Biol. Inorg. Chem.* **2015**, *20*, 373

[52] Raaijmakers, H. et al. *J. Biol. Inorg. Chem.* **2006**, *11*, 849

[53] Mota, S. et al. *J. Biol. Inorg. Chem.* **2011**, *16*, 1255

[54] Maia, L.; Fonseca, I.; Moura, I.; Moura, J. *J. Am. Chem. Soc.* **2016**, *138*, 8834

[55] Robinson, W.; Bassegoda, A.; Reisner, E.; Hirst, J. *J. Am. Chem. Soc.* **2017**, *just accepted*,

[56] Boyington, J.; Gladyshev, V.; Khangulov, S.; Stadtman, T.; Sun, P. *Science* **1997**, *275*, 1305

[57] Eichinger, M. *Phd-Thesis* **1999**, München, Germany

[58] Grigorieff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. *J. Mol. Biol.* **1996**, *259*, 393

[59] Essen, L.-O.; Siegert, R.; Lehmann, W. D.; Oesterhelt, D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11673

[60] Patzelt, H.; Simon, B.; terLaak, A.; Kessler, B.; Kuhne, R.; Schmieder, P.; Oesterhelt, D.; Oschkinat, H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9765

[61] Kandt, C.; Schlitter, J.; Gerwert, K. *Biophys. J.* **2004**, *86*, 705

[62] Grudinin, S.; Bu¨ldt, G.; Gordeliy, V.; Baumgaertner, A. *Biophys. J.* **2005**, *88*, 3252

[63] Mroginski, M. A.; Kaminski, S.; von Stetten, D.; Ringsdorf, S.; Gärtner, W.; Essen, L.-O.; Hildebrandt, P. *Phys. Chem. B* **2011**, *115*, 1220

[64] Daminelli et al. **unpublished (at the time of writing this thesis)**

[65] Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2002**, *83*, 1281

[66] Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 453

[67] Babitzki, G.; Denschlag, R.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10483

[68] Yang, X.; Kuk, J.; Moffat, K. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15639

[69] Salewski, J.; Escobar, F. Velazquez; Kaminski, S.; von Stetten, D.; Keidel, A.; Rippers, Y.; Michael, N.; Scheerer, P.; Piwowarski, P.; Bartl, F.; Frankenberg-Dinkel, N.; ringsdorf, S.; Gärtner, W.; Lamparter, T.; Mroginski, M. Andrea; Hildebrandt, P. *J. Biol. Chem.* **2013**, *288*, 16800

[70] Yang, X.; Ren, Z.; Kuk, J.; Moffat, K. *Nature* **2011**, *479*, 428

[71] Leopoldini M.; Chiodo S.; Toscano M.; Russo N. *Chem. Eur. J.* **2008**, *14*, 8674

[72] Niks, D.; Duvvuru, J.; Escalona, M.; Hille, R. *J. Biol. Chem.* **2016**, *291*, 1162

[73] Wahlefeld, S.; Belger, D.; Zegber, I.; Hildebrandt, P.; Leimkühler, S; Hartmann, T.; Dufus, B.; Schulzke, C.; Nowak, R.; Mroginski, M. et al. **Unpublished**; Vibrational spectroscopic data concerning *Rc*FDH can also be found in: Wahlefeld, S. *Phd-Thesis* **2018**, Berlin, Germany

[74] Schrödinger, E. *Ann. Phys.* **1926**, *79*, 361

[75] Krese, G.; Hafner *J. Phys . Rev. B* **1993**, *47*, 558

[76] Ryckaert, J.; Bellemans, A. *Faraday Discuss.* **1978**, *66*, 95

[77] Rahman, A.; Stillinger, F. *J. Chem. Phys.* **1971**, *55*, 3336

[78] Tersoff, J. *Phys. Rev. B* **1989**, *39*, 5566

[79] Cohen, R.; Krakauer, H. *Phys. Rev. B* **1990**, *42*, 6416

[80] Catlow, C.; Price, G. *Nature* **1990**, *347*, 243

[81] Karplus, M.; McCammon, J. *Nat. Struct, Biol.* **2002**, *9*, 788

[82] Van Gunsteren, W.; Mark, A. *Eur. J. Biochem.* **1992**, *204*, 947

[83] Tielemann, D.; Marrink, S.; Berendsen, H. *BBA* **1997**, *1331*, 235

[84] Mroginski, M. *"Vibrational Spectroscopy and more (calculations) 8. Modelling biomolecules"* **2015/16**, *Presentation/Seminar*, TU Berlin

[85] MacKerell, Jr, A.; Bashford, D.; Bellott, M. et al. *J. Phys. Chem. B* **1998**, *102*, 3586

[86] Mroginski, M. *"Vibrational Spectroscopy and more (calculations) 9. Simulating dynamics"* **2015/16**, *Presentation/Seminar*, TU Berlin

[87] Allen, M.; Tildesley, D., *"Computer simulation of liquids"*, Oxford University Press USA, **1990**

[88] Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684

[89] Andersen, H. *J. Chem. Phys.* **1980**, *72*, 2384

[90] Frenkel, D.; Smit, B., *"Understanding molecular simulation: from algorithms to applications"*, Academic Pres, **2002**

[91] Brooks III, C.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312

[92] Berkowitz, M.; McCammon, J. *Chem. Phys. Lett.* **1982**, *90*, 215

[93] Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924

[94] Leach, A., *"Moldecular Modelling: principles and applications"*, Addison-Esley Longman Ltd., **2001**

[95] Ewald, P. *Ann. Phys.* **1921**, *64*, 300

[96] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089

[97] Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7694

[98] Mulliken, R. *J. Chem. Phys.* **1955**, *23*, 1833

[99] Ryckaert, J.; Ciccotti, G.; Berendes, H. *J. Comput. Phys.* **1977**, *23*, 327

[100] Barth, E.; Kuczera, K.; Leimkühler, B.; Skeel, R. *J. Comput. Chem.* **1995**, *16*, 1192

[101] Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. . *Chem. Phys.* **1983**, *79*, 926

[102] *"Root Mean Square"*, https://en.m.wikipedia.org/wiki/Root_mean_square, **19.10.2018**

[103] *"A Dictionary of Physics"*, 6, Oxford University Press, **2009**

[104] *"Root Mean Square Deviation"*, https://en.m.wikipedia.org/wiki/Root-mean-square_deviation, **19.10.2018**

[105] Hyndman, R. J.; Koehler, A. B. *Int. J. Forecasting* **2006**, *22*, 679

[106] *"Root Mean Square Deviation of Atomic Positions"*, https://en.m.wikipedia.org/wiki/Root-mean-square_deviation_of_atomic_positions, **19.10.2018**

[107] Born, M.; Oppenheimer, J. *Ann. Phys.* **1927**, *84*, 457

[108] Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, 864

[109] MacDonald, J. *Phys. Rev.* **1933**, *43*, 830

[110] Kohn, W.; Sham, L. et al. *Phys. Rev. A* **1965**, *140*, 1133

[111] Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B* **1988**, *37*, 785

[112] Becke, A. *Phys. Rev. A* **1988**, *38*, 3098

[113] Slater, J. *Phys. Rev.* **1951**, *81*, 385

[114] Dirac, P. *Proc. Royal Soc. (London) A* **1929**, *123*, 714

[115] Voskov, S.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200

[116] Mroginski, M. *"Vibrational Spectroscopy and more (calculations) 6. "crash course" to quantum chemistry"* **2015/16**, *Presentation/Seminar*, TU Berlin

[117] Elstner, M. et al. *Phys. Rev. B* **1998**, *58*, 7260

[118] Levine, I., *"Quantum Chemistry"*, Prentice-Hall, 4, **1991**

[119] Hellmann, H. *J. Chem. Phys.* **1935**, *3*, 61

[120] Louie, S.; Froyen, S.; Cohen, M. *Phys. Rev. B* **1982**, *26*, 1738

[121] Reis, C.; Pacheco, J.; Martins, J. *Phys. Rev. B* **2003**, *68*, 155111

[122] Harrison, W., *"Pseudopotentials in the theory of metals, Frontiers in physics"*, University of Virginia, **1966**

[123] Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117

[124] Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239

[125] Pascual-Ahuir, J.; Silla, E.; Tunon, I. *J. Comp. Chem.* **1994**, *15*, 1127

[126] Fogolari, F.; Brigo, A.; Molinari, H. *J. Mol. Rec.* **2002**, *15*, 377

[127] Koehl, P. *Curr. Opin. Struct. Biol.* **2006**, *16*, 142

[128] Onufriev, A.; Case, D.; Bashford, D. *J. Comp. Chem.* **2002**, *23*, 1297

[129] Still, W.; Tempctyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127

[130] Dokolyan, N.; Fuxreiter, M. *"Computational Approaches to Protein Dynamics – From Quantum to Coarse-Grained Methods"* **2014**, *1*, 15

[131] Wheeler, R.; Spellmeyer, D. *"Annual Reports in Computational Chemistry"* **2010**, *6*, 121

[132] Dao-pin, S.; Anderson, D.; Baase, W.; Dahlquist, F.; Matthews, B. *Biochem.* **1991**, *30*, 11521

[133] Caputo, G.; London, E. *Biochem.* **2003**, *42*, 3275

[134] Schaefer, M.; van Vlijmen, H.; Karplus, M. *"Advances in Protein Chemistry"* **1998**, *51*, 1

[135] *Lecture Friedrich-Schiller-Universität Jena "Kapitel 7: Lösungsmitteleinfluss"*, https://www.ipc.uni-jena.de/downloads/Lehre/VF_ThCh/SS/vl_vf_kap7.pdf, **11.4.2017**

[136] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999

[137] Cerqueira N.; Fernandes P.; Eriksson L.; Ramos M. *Biophys. J.* **2006**, *90*, 2109

[138] Cerqueira N.; Fernandes P.; Eriksson L.; Ramos M. *J. Mol. Struct. Theochem.* **2004**, *709*, 53

[139] Himo F. *Theor. Chem. Acc.* **2006**, *116*, 232

[140] von Hippel, A., *"Dieelectric Materials and Applications"*, Artech Hous, **1954**

[141] Yamane, T.; Inoue, Y.; Sakurai, M. *Chem. Phys. Let.* **1998**, *291*, 137

[142] *"The Polarizable Continuum Model (PCM)"*, https://www.cup.uni-muenchen.de/oc/zipse/teaching/computational-chemistry-2/topics/the-polarizable-continuum-model-pcm/, **16.07.2019**

[143] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227

[144] Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198

[145] Thery, V. et al. *J. Comp. Chem.* **1994**, *15*, 269

[146] Gao, J.; Amara, P.; Alhambra, C.; Field, M. *J. Phys. Chem. A* **1998**, *102*, 4714

[147] Murphy, R.; Philipp, D.; Firesner, R. *J. Comp. Chem.* **2000**, *21*, 1442

[148] Field, M.; Bash, P.; Karplus, M. *J. Comp. Chem.* **1990**, *11*, 700

[149] Singh, U.; Kollman, P. *J. Comp. Chem.* **1986**, *7*, 718

[150] Brooks, B.; Bruccoöero. R.; Olafson, B. *J. Comput. Chem.* **1983**, *4*, 187

[151] Karplus et al. *Phys. Chem. B* **1998**, *102*, 3586

[152] Cornell, W. et al. *J. Am. Chem. Soc.* **1995**, *117*, 5179

[153] Frisch, M. et al., *Gaussian 09, Revision A.02*, Gaussian, Inc., Wallingford CT, **2016**

[154] Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165

[155] *"Chemshell"*, www.chemshell.org, **5.5.2017**

[156] Siebert, F.; Hildebrandt, P., *"Vibrational Spectroscopy in Life Science"*, Wiley, 1, **2008**

[157] Hildebrandt, P. *"Vibrational Spectroscopy and related topics – 1.Theory of vibrational spectroscopy"* **2015/16**, *Presentation/Seminar*, TU Berlin

[158] Mroginski, M. *"Vibrational Spectroscopy and more (calculations) 7. Computation of observables"* **2015/16**, *Presentation/Seminar*, TU Berlin

[159] Raman, C.; Krishnan, K. *Nature* **1928**, *121*, 501

[160] Czerunszewicz, R.; Spiro, T., *"Inorganic Electronic Structure and Spectroscopy, Vol. I: Methodology"*, Wiley, **1999**

[161] Wilson, E.; Decius, J.; Cross, P., *"Molecular vibrations: the theory of infrared and Raman vibrational spectra"*, Dover Publications, **1980**

[162] Smit, K. *Phd thesis* **1992,** Duisburg, Germany

[163] Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. *J. Am. Chem. Soc.* **1979**, *101*, 2550

[164] Heisenberg, W. *Zeitschrift für Physik* **1927**, *43*, 172

[165] Mroginski, M.; von Stetten, D.; Escobar, F.; Strauss, H.; Kaminski, S.; Scheerer, P.; Gunther, M.; Murgida, D.; Schmieder, P.; Bongards, C.; Gartner, W.; Mailliet, J.; Hughes, J.; Essen, L.; Hildebrandt, P. *Biophys. J.* **2009**, *96*, 4153

[166] Ibragimova, G.; Wade, R. *Biophys. J.* **1998**, *74*, 2906

[167] McGuffin, L.; Bryson, K.; Jones, D. *Bioinformatics* **2000**, *16*, 404

[168] Brünger, A.; Karplus, M. *Proteins* **1988**, *4*, 148

[169] Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comp. Chem.* **2005**, *26*, 1781

[170] van Gunsteren, W.; Berendsen, H. *Mol. Phys.* **1977**, *34*, 1311

[171] Martyna, G.; Tobias, D.; Klein, M. *J. Chem. Phys.* **1994**, *101*, 4177

[172] Feller, S.; Zhang, Y.; Pastor, R.; Brooks, B. *J. Chem. Phys.* **1995**, *103*, 4613

[173] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33

[174] Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452

[175] Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 93

[176] Sherwood, P. et al. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1

[177] Billeter, S.; Turner, A.; Thiel, W. *Phys. Chem. Chem. Phys* **2000**, *2*, 2177

[178] Hildebrandt, P.; Hoffmann, A.; Lindemann, P.; Heibel, G.; Braslavsky, S. E.; Schaffner, K.; Schrader, B. *Biochem.* **1992**, *31*, 7957

[179] Fodor, S.; Lagarias, J.; Mathies, R. *Biochem.* **1990**, *29*, 11141

[180] *"Methylation Kit"*, http://www.jenabioscience.com,

[181] Schaftenaar, G.; Noordik, J. *J. Comput.-Aided Mol. Design* **2000**, *14*, 123

[182] Peters, J.; Lanzilotta, W.; Lemon, B.; Seefeldt, L. *Science* **1998**, *282*, 1853

[183] Sazanov, L.; Hinchliffe, P. *Science* **2006**, *311*, 1430

[184] Dennington, R.; Keith, T.; Millam, J., *"GaussView, Version 6"*, Semichem Inc., Shawnee Mission, KS, **2016**

[185] Ditchfield, R.; Hehre, W.; Pople, J. *J. Chem. Phys.* **1971**, *54*, 724

[186] Francl, M.; Pietro, W.; Hehre, W.; Binkley, J.; DeFrees, D.; Pople, J. *J. Chem. Phys.* **1982**, *77*, 3654

[187] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297

[188] Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123

[189] Porcher, J.; Fogeron, T.; Gomez-Mingot, M.; Derat, E.; Chamoreau, M.; Li, Y.; Fontecave, M. *Angew. Chem. Int. Ed.* **2015**, *54*, 14090

[190] Ochterski, J., *"Thermochemistry in Gaussian"*, help@gaussian.com, ©2000, Gaussian, Inc., **June 2, 2000** (https://gaussian.com/thermo/, 16.07.2019)

[191] *"ChemDraw", PerkinElmer Informatics*, https://www.cambridgesoft.com/support/DesktopSupport/KnowledgeBase/FAQ/details/Default.aspx?TechNote=3411,

[192] *"Norddeutscher Verbund für Hoch- und Höchstleistungsrechne (HLRN)"*, www.hlrn.de, **15.01.2019**

[193] *"Rechencluster - Technische Universität Berlin, Insitut für Mathematik"*, www.math.tu-berlin.de/iuk/forschungsrechnerbereich/service/cluster_nutzung/, **15.01.2019**

[194] Cornilescu, G.; Ulijasz, A.; Cornilescu, C.; Markley, J.;Vierstra, R. *J. Mol. Biol.* **2008**, *383*, 403

[195] Ulijasz, A.; Cornilescu, G.; Cornilescu, C.; Zhang, J.; Rivera, M.; Markley, J.; Vierstra, R. *Nature* **2010**, *463*, 250

[196] Dufus, B. et al. **unpublished (at the time of writing this thesis)**

[197] Rivas, M.; Gonzalez, P.; Brondino, C.; Moura, J.; Moura, I. *J. Inorg. Biochem.* **2007**, *101*, 1617

[198] Dwyer, J.; Gittis, A.; Karp, D.; Lattman, E.; Spencer, D.; Stites, W.; Garcia-Moreno, B. *Biophys. J.* **2000**, *79*, 1610

[199] Warshell, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211

[200] Mroginski, M. et al. **unpublished (at the time of writing this thesis)**

[201] Lever, A. *Inorg. Chem.* **1990**, *29*, 1271

[202] Hansch, C.; Leo, A.; Taft, R. *Chem. Rev.* **1991**, *91*, 165

[203] Popov, V.; Lamzin, V. *Biochem. J.* **1994**, *301*, 625

[204] Blanchard, J.; Cleland, W. *Biochem.* **1980**, *19*, 3543

[205] Joshi, H.; Cooney, J.; Inscore, F.; Gruhn, N.; Lichtenberger, D.; Enemark, J. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3719